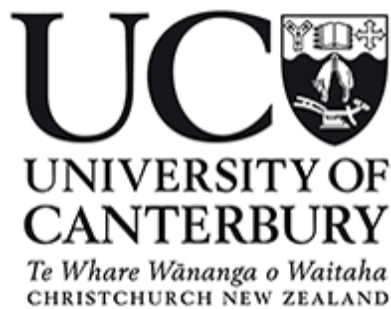


Department of Physics and Astronomy, University of Canterbury, Private Bag 4800  
Christchurch New Zealand

# Kinect Surface Tracking

*James T Eagle (BSc)*

A thesis submitted in partial fulfilment of the requirements for the degree of Master of  
Science in the University of Canterbury



Supervisor: Dr. Steven Marsh

Co-supervisors: Associate Professor Dr. Juergen Meyer, Dr. Adrian Clark

2014



# Abstract

---

A cost effective method for evaluating patient motion during external beam radiotherapy has been developed. This system consists of a 3D depth camera and tracking software. The system was tested using a range of custom built phantoms and motion platforms. Kinect v1 and v2 sensors were chosen for this project due to their low cost and high resolution as compared to other depth cameras. Various tracking algorithms were evaluated to determine the optimal method for the system.

Testing determined that the Kinect v2 was the more suitable sensor primarily due to the significantly lower depth variation of 0.6mm as compared to the Kinect v1 of 2.4mm. An analysis of tracking methods was performed and concluded that the least squares method provides the most accurate tracking in the clinical environment. This method is able to determine an object's actual position accurate to 1.5mm 66% of the time.

Testing in a clinical environment on a volunteer demonstrated that the system was able to detect a broad range of clinically relevant motions. Movements ranging in amplitude from shallow breathing with a maximum amplitude of 8mm to large displacements of 80mm were successfully detected. Further evaluation demonstrated that the depth movements of 1mm or greater are able to be observed however, only horizontal and vertical movements greater than 5mm were observed. These results suggest that the system could be used in treatments that have approximately 5mm tolerance for patient movement. Further work to improve accuracy is required so that the system is able to detect very small movements in order for use in highly conformal methods for example SBRT.

# Acknowledgements

---

Firstly I would like to thank my Supervisor Dr Juergen Meyer for all his support and contribution to my thesis. I thoroughly enjoyed working with him in Seattle and am grateful the opportunity to travel over there for part of this work.

I would like to thank the University of Washington Medical Centre, Radiation Oncology Medical Physics department for allowing me to use their facilities and opportunity to work in their department.

I would like to thank The University of Canterbury Physics and Astronomy department for their support. Especially Dr Mike Reid for his input and support through my thesis and Dr Steve Marsh for his support and help proof reading my thesis. I would also like to thank the workshop staff in particular Wayne Smith and Robert Thirkettle for their help in production of components for this thesis.

I would like to thank St Georges Radiation Oncology department for access and use of their facilities.

Finally I would like to thank my family and friends. Epically, Mum and Dad for being there and supporting me throughout the entire thesis and University studies.

I wish to gratefully acknowledge the financial assistance from the University of Canterbury Physics and Astronomy department.

# Contents

---

Kinect Surface Tracking .....	i
Abstract.....	iv
Acknowledgements.....	vi
Contents.....	viii
List of Figures.....	xii
List of Tables.....	xiv
Chapter 1: Introduction .....	1
1.1 Motivation.....	3
Chapter 2: Background.....	6
2.1 Ionising radiation.....	7
2.1.1 Production .....	10
2.1.2 Linear accelerators .....	15
2.1.3 Biological interactions .....	17
2.2 Cancerous tissue .....	17
2.2.1 Imaging and detection.....	19
2.2.2 Computer Tomography.....	20
2.3 Radiotherapy.....	22
2.3.1 Beam shaping .....	24
2.3.2 Target volumes.....	25
2.4 Radiotherapy Modalities .....	27
2.5 Patient motion management.....	30
2.5.1 Patient Positioning.....	32
2.5.2 Patient motion management.....	33
2.5.3 Surface tracking.....	34
2.6 Previous work.....	34



Chapter 3: Principles of Depth Cameras .....	37
3.1 Depth Cameras .....	37
3.1.1 Time of Flight.....	38
3.1.2 Time of flight camera .....	39
3.1.3 Triangulation.....	42
3.2 Kinect for Xbox™ 360 .....	46
3.3 Kinect for Windows v2 .....	49
Chapter 4: Tracking Methods.....	51
4.1 Mean Shift.....	53
4.2 Camshift .....	56
4.3 Scale-invariant feature transform (SIFT) .....	60
4.3.1 Speed Up Robust Features (SURF).....	62
4.4 Least Squares.....	63
4.5 Iterative closest point (ICP).....	65
4.6 Kinect Fusion.....	67
Chapter 5: Software .....	70
5.1 Development Environment .....	71
5.1.1 Programing libraries and Software Development Kit (SDK) .....	71
5.2 Graphical User Interface (GUI).....	73
5.3 Parallelisation .....	77
5.4 Coordinate systems .....	78
5.4.1 Kinect Sensors co-ordinates.....	79
5.4.2 LINAC couch coordinate system.....	80
5.5 Kinect to LINAC coordinate transformation. ....	81
Chapter 6: Characterisation.....	86
6.1 Sensor placement and setup methodology .....	86
6.2 Hardware .....	87
6.2.1 Kinect Long term stability .....	88

6.2.2	Kinect Depth resolution .....	94
6.2.3	Vertical and horizontal resolution .....	98
6.2.4	Kinect subsampling.....	99
6.3	Software .....	100
6.3.1	Kinect noise and averaging.....	100
6.3.2	Iso-Mapping .....	102
6.4	Tracking software testing .....	104
6.4.1	Motion tracking and motion phantom.....	105
6.4.2	Camshift .....	107
6.4.3	Kinect Fusion and ICP .....	109
6.4.4	Least squares .....	111
6.4.5	Surf.....	113
6.5	Summary .....	115
Chapter 7: Clinical Testing.....		118
7.1	Clinical Setup .....	119
7.2	Position testing.....	122
7.3	Volunteer testing.....	125
7.3.1	Baseline.....	126
7.3.2	Normal breathing.....	127
7.3.3	Heavy breathing.....	128
7.3.4	Coughing.....	130
7.3.5	Looking around.....	131
7.3.6	Moving their backside .....	131
7.3.7	Talking .....	132
7.3.8	Arm movements.....	133
Chapter 8: Discussion and Conclusion.....		135
8.1	Hardware .....	136
8.2	Software .....	137

8.2.1	Tracking software.....	137
8.2.2	Position and Clinical testing .....	138
8.3	Future work.....	140
	Bibliography.....	142

# List of Figures

---

Figure 2.1: Illustration of the interactions between photons and electrons .....	7
Figure 2.2: High Energy photon interactions with lead.....	9
Figure 2.3: Illustrations of a cobalt treatment machine.....	10
Figure 2.4: Illustration of X-ray production].....	12
Figure 2.5: X-ray spectrum from a silver target.....	14
Figure 2.6: Diagram of a traveling wave LINAC.....	15
Figure 2.7: Diagram of a Linear accelerator] .....	16
Figure 2.8: Screenshots from an IMRT Prostate plan.....	23
Figure 2.9: Image of a Varian 120-leaf MLC.....	25
Figure 2.10: Diagram illustrating the main radiotherapy planning volumes.....	26
Figure 2.11: Illustration of patient workflow in a Radiation Oncology .....	31
Figure 2.12: Screenshot of Talbot's system .....	36
Figure 3.1: Diagram of a RF modulated TOF camera.....	40
Figure 3.3: Illustration of a triangulation camera.....	42
Figure 3.4: Illustration of parallel line encoding .....	44
Figure 3.5: Image of the Kinect for Xbox 360 structured light pattern].....	46
Figure 3.6: Image of the Kinect v1 and Kinect v2 sensors.....	48
Figure 4.1: Diagram of the Kinect coordinate system.....	51
Figure 4.2: Illustration of object movement in depth images.....	52
Figure 4.3: Flow diagram of Camshift Tracking method.....	56

Figure 4.4: Illustration of the HSV colour space .....	57
Figure 4.5: Point cloud inside extended point cloud .....	64
Figure 4.6: Kinect Fusion reconstruction and pipeline.....	68
Figure 5.1: Graph component of the GUI.....	74
Figure 5.2: Screenshot of the current GUI Design .....	74
Figure 5.3: Control section of GUI .....	75
Figure 5.4: Screenshot GUI Design demonstrating the colour warning system .....	77
Figure 5.5: Illustration of the Kinect coordinate system.....	79
Figure 5.6: Illustration of trigonometric calculation .....	80
Figure 5.7: Model of LINAC and couch movements and coordinates.....	81
Figure 5.8: Image of the 3D model of the Isocentre cube.....	83
Figure 5.9: Image of the Isocentre Cubes .....	82
Figure 6.1: Electronic stability of the Kinect v1 and v2 .....	89
Figure 6.2: Image showing the Kinect stability testing setup .....	89
Figure 6.3: Electronic stability of the Kinect v1 and Kinect v2 .....	91
Figure 6.4: Histogram of Electronic stability over period of time Kinect v2 .....	92
Figure 6.5: Histogram of Electronic stability over period of time Kinect v1 .....	92
Figure 6.6: Difference graph between Kinect v1 and the correct position.....	96
Figure 6.7: Difference graph between Kinect v2 and the correct position .....	97
Figure 6.8: Illustration demonstrating a divergent FOV .....	98
Figure 6.9: Image demonstrating the discreteness observed.....	99
Figure 6.10 Graph demonstrating the effect of averaging on total noise .....	101
Figure 6.11: Graph displaying the error as a function of rotation.....	103
Figure 6.12: Image of the Motion breathing platform .....	105

Figure 6.13: Graph showing Camshift tracking data .....	107
Figure 6.14: Graphs showing the difference between the Camshift tracking data.....	108
Figure 6.15: Graph showing Kinect Fusion tracking data .....	110
Figure 6.16: Graphs showing the difference between the Kinect Fusion tracking data ....	111
Figure 6.17: Graph showing Least squares tracking data .....	112
Figure 6.18: Graphs showing the difference between the Least squares tracking data...	113
Figure 6.19: Graph showing Surf tracking data .....	114
Figure 6.20: Graphs showing the difference between the Surf tracking data.....	114
Figure 7.1: Image of camera placement and setup at the proton therapy centre.....	120
Figure 7.2: Testing of the system with movements in the lateral direction .....	122
Figure 7.3: Testing of the system with movements in the vertical direction .....	123
Figure 7.4: Testing of the system with movements in the longitudinal direction.....	123
Figure 7.5: Base line movements of the volunteer over 25 seconds. ....	126
Figure 7.6: Normal breathing of the volunteer over 37 seconds.....	128
Figure 7.7: Heavy breathing of the volunteer over 65 seconds .....	129
Figure 7.8: Coughing testing of the volunteer.....	130
Figure 7.9: Volunteer looking around of the volunteer for a duration of 40 seconds....	131
Figure 7.10: Volunteer readjusting backside.....	132
Figure 7.11: Volunteer talking for a time of 75 seconds.....	133
Figure 7.12: Volunteer moving their arms for 33 seconds.....	134

## List of Tables

Table 1: Kinect hardware comparison .....	115
Table 2: Tracking methods comparison.....	116
Table 3: Volunteer testing results .....	135
Table 4: Tracking methods comparison.....	138
Table 5: Volunteer testing results .....	139

# Chapter 1: Introduction

---

Cancer is responsible for 29.4% of all deaths in New Zealand [1], and in 2012 was responsible for 8.2 million deaths worldwide [2]. The International Agency for Research on Cancer (IARC), estimated that there were 14 million new cases of cancer in 2012, and within the next 20 years this figure is expected to rise to 22 million annually [3]. It was estimated in 2010 that the total economic cost of cancer was \$1.16 trillion US dollars worldwide [3].



Current approaches for cancer control and treatment includes surgery, chemotherapy, and radiotherapy [4]. They can be used individually or in conjunction with each other. The role radiotherapy plays in treatment of cancer is becoming ever more important due to the increase number of cancer cases each year. Research undertaken by G Delaney et al. [5] estimated that between 51.7% and 53.1% of all cancer would benefit from radiotherapy. Radiotherapy is either used for local control of a tumour, when the tumour is localised to a particular region of the patient's body, or in palliative treatments. Radiation therapy employs high energy ionising radiation to kill cancer cells. The delivery method utilised to irradiate a patient depends on the type and location of the patient's cancer. This can either be with external or internal sources. This research is only applicable to external beam sources and therefore, further discussion in this work is only concerned with external beam sources. External beam radiation treatment using photons is the most commonly available external treatment worldwide [6]. External beams are generated by either radioactive decay from an unstable radionuclide source, or from a particle accelerator. The most common method used in New Zealand and worldwide is by use of linear accelerators (LINACS).

Radiation is used to kill cells by damaging a cell's DNA resulting in the cells death. Ionising radiation's ability to kill cells and tissue is used to treat cancer by killing the cancerous cells. However, radiation does not distinguish between healthy tissue and cancerous tissue and therefore treatments have to be planned and delivered carefully. All tissue which a beam passes through will receive a dose of radiation. The ratio between the killing of tumour cells (therapeutic effect) and the killing of healthy cells (toxicity) is known as the Therapeutic Ratio. Increasing the therapeutic ratio will result in improved outcomes for a patient. To limit the damage to healthy tissue and increase the damage to tumour cells, multiple beams can be used alongside beam shaping.

## 1.1 Motivation

Prior to a radiotherapy treatment, patients are setup in a predetermined position to best enable the radiation to be correctly delivered to the tumour. Throughout the treatment patients are required to remain as still as possible as any movement will cause them to move out of the correct position, therefore moving the tumour out of the radiation beam and healthy tissue into the beam. Several methods are employed to assist the patient to remain as immobile as possible for the entire duration of the treatment. Various constraints are used to immobilise a patient. One common method for example for head and neck treatments are facial masks. They are plastic masks moulded to a patients head and are attached to a table immobilising their head. These can be quite successful in head or neck treatments, however they are less useful when treating mobile tumour sites around the lungs. For thoracic treatments full or half body bags may be used to immobilise the patient but they do not suppress breathing motion. One possible solution is the development of a patient surface monitoring systems.

In recent years a few commercial systems have been developed to address patient position and breathing monitoring issue. These systems are not yet common in radiation oncology departments. In order for a patient positioning monitoring system to be effective, the system must meet certain specifications. Such as detection of small movements, produces no ionising radiation, rapid motion detection and measurements must be in usable units and coordinates. Detection of small movements is required as this is to determine if a patient remains in the correct position for small radiation fields used in modern radiotherapy. Ionising radiation should not be used in order to reduce the negative impact that any unnecessary radiation will have on healthy tissue. Detection of motion must be rapid so that should a patient move out of the correct position during treatment, the radiation can be stopped promptly. It should also be capable of detecting and measuring real distances in all three dimensions (3D) as all modern treatments are treating a 3D tumour volume.

In summary there are a growing number of cancer cases each year of which approximately 52% would benefit from radiotherapy. In order for successful treatment, patients must remain still throughout treatment. However, currently there are a lack of cost effective patient position monitoring systems. The purpose of this research is to determine the feasibility of a cost effect system using consumer hardware to address these issues.

This thesis is divided into eight chapters the next chapter introduces the background material of radiotherapy detailing the need for a surface tracking system. Chapter 3 introduces possible cameras which a tracking system could utilise, and the physics behind these cameras and why depth camera are needed for this research. Chapter 4 discusses the need for a tracking software integration in the system, and various tracking methods which are tested throughout this research. Chapter 5 introduces the original work done by the author on development and design of the system and software. Chapter 6 presents measurements and characterisation testing preformed to test the various components of the system. Chapter 7 demonstrates and tests the feasibility of the system in a clinical environment on phantoms and volunteers. Chapter 8 concludes the thesis discussing the results found and possible future work proceeding from this thesis.

# Chapter 2: Background

---

This chapter introduces background material relevant to this thesis. The chapter starts with a broad introduction to radiation, how it is produced, and how it is used to treat cancer. The chapter then discusses what cancer is and how it is detected. This is followed by a section on radiation therapy and the modalities used and then concludes with a section on patient motion management.

## 2.1 Ionising radiation

Ionising radiation is high energy radiation which when interacting with an atom will cause an electron to be freed, resulting in the atom to be ionised [7]. Ionising radiation can be split into two primary groups for medical physics: directly ionising and indirectly ionising [7]. Directly ionising radiation interacts with atoms through the Coulomb force requiring the radiation to be charged. Alpha particles, beta particles and heavy ions are all directly ionising radiation. Indirectly ionising radiation has no charge and therefore must ionise atoms through secondary effects. Non-directly ionising radiation consists of high frequency photons and neutrons. Since Radiotherapy throughout the world primarily uses high energy photons for treatment [6], the following section refers to this form of ionising radiation. Photons are electromagnetic waves consisting of high energy ultraviolet, X-rays or gamma rays. There are three interactions with matter that a high energy photon can have which results in ionising radiation. These types are: the Photoelectric effect, Compton scattering and Pair production [8]. These interactions are illustrated in Figure 2.1. If a photon of sufficient energy collides with an electron in an atom, the electron can gain sufficient energy to escape the atom, and will result in the photon being absorbed and an electron being emitted. [8] This is called the Photoelectric effect and dominates photon interactions with electrons at low energy.

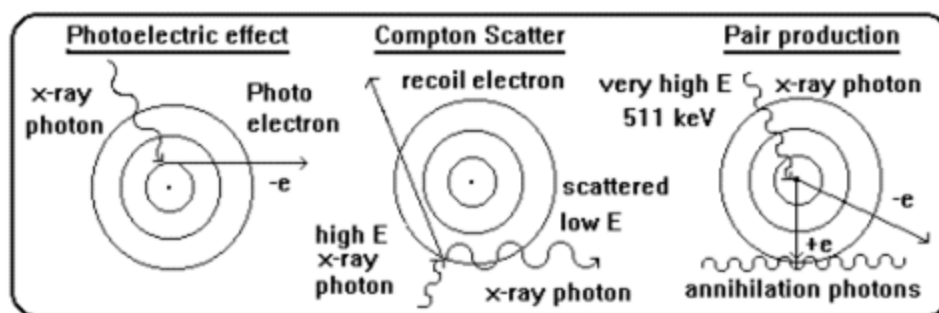


Figure 2.1: Illustration of the interactions between photons and electrons [39]

Compton Scattering is an inelastic scattering event of a photon and an electron where a portion of the photon energy is delivered to the electron resulting in its' ionisation. [8] The resulting photon emitted after the collision will have a reduced energy and a different direction. Pair production only occurs at high energy and dominates above 5MeV. When a photon with more than 1.022MeV energy collides with an atom, an electron and positron pair can be produced. This occurs due to the photons energy being converted to mass, this is governed by

$$E = mc^2.$$

( 2.1)

Where  $E$  is the inbound photon energy,  $m$  is the combined mass of the electron and positron and  $c$  is the speed of light. The rest energy of an electron and positron is 0.511MeV. Therefore the initial condition for pair production to occur is given by the combined rest energy of the electron positron pair. The interactions between photons of various energies and lead is illustrated in Figure 2.2. This shows that the photoelectric effect dominates the total interactions at low energies and pair production dominates the total interactions at high energy.

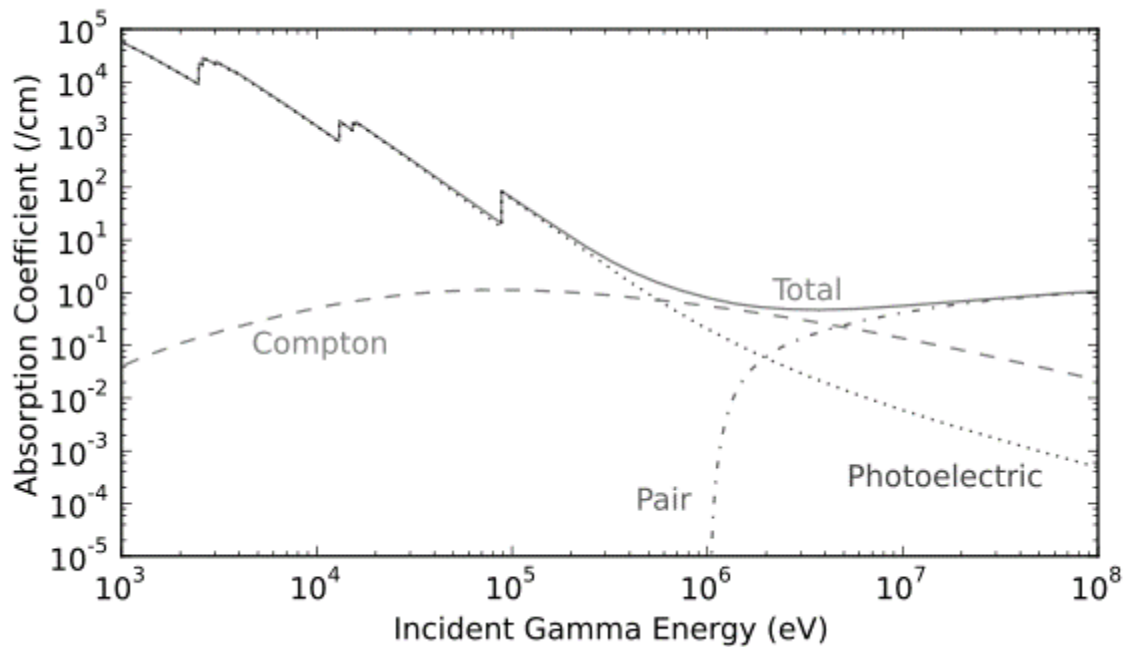
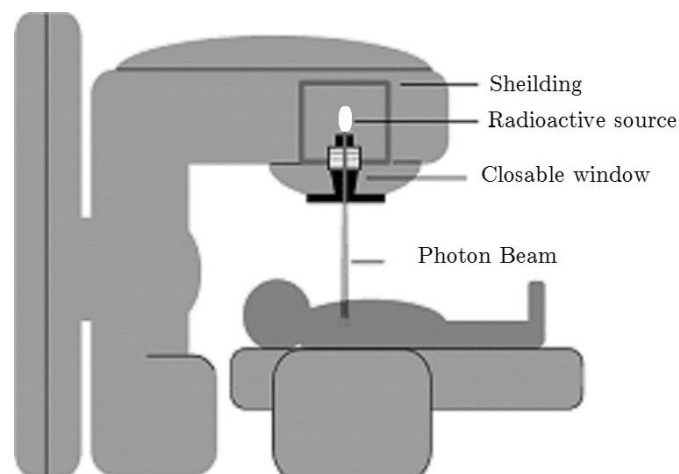


Figure 2.2: High Energy photon interactions with lead, demonstrating the three primary interactions and the energy dependence these interactions have [21]



### 2.1.1 Production

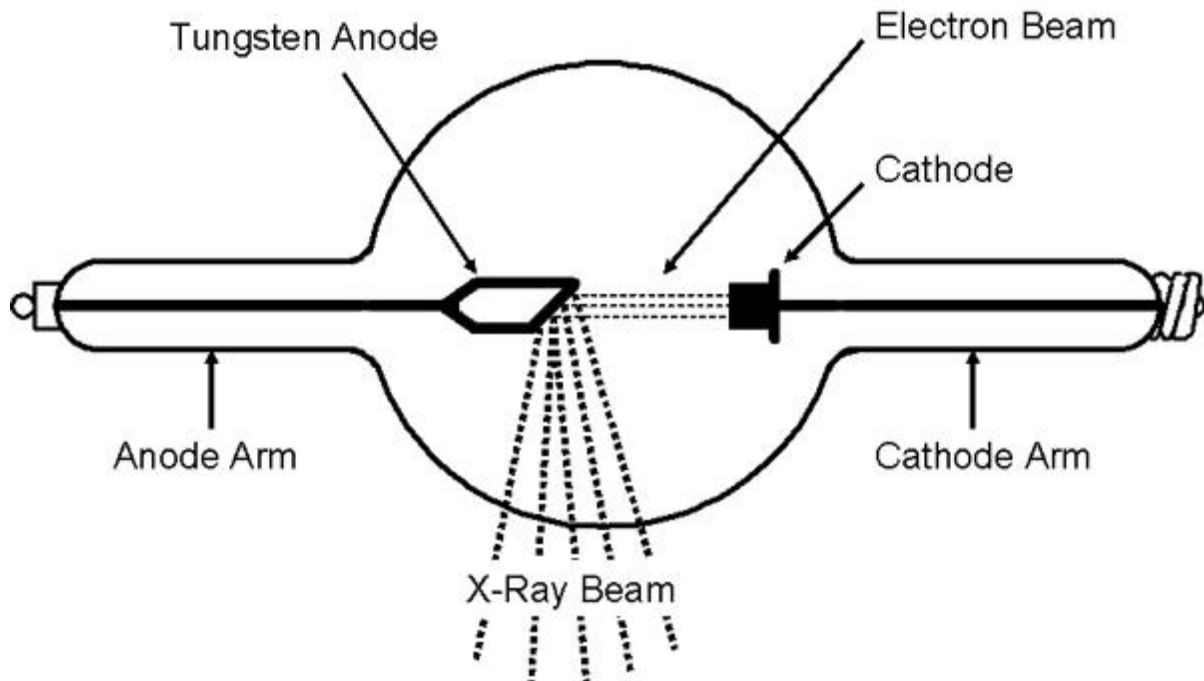
High energy photons are termed as either X-rays or gamma rays. The difference between the two types of radiation is entirely historical [9], as the resulting photons are identical. X-rays are defined as photons emitted from electrons, and Gamma rays are defined as being produced from an atom's nucleus. For a gamma ray to be produced a radioactive isotope can decay emitting a range of particles including gamma, beta and alpha radiation. The production of gamma radiation results from an unstable nucleus dropping from a high energy state to a low energy state. This regular decay from unstable isotopes can be



*Figure 2.3: Illustrations of a cobalt treatment machine, demonstrating the machine treating a patient on the treatment couch, adapted from [22]*

used to generate a photon beam for clinical use. A radioactive source, often cobalt-60 is placed inside a treatment machine. This is illustrated in Figure 2.3 where a patient is placed on a treatment couch receiving a photon beam from a radioactive source.

As defined earlier, Gamma rays can result from the nucleus of unstable atoms and X-rays are produced from orbitals changing from a high energy state to a low energy state. X-rays are produced by firing electrons into atoms with a high proton number or Z number. The simple case of X-ray production with an X-ray tube is illustrated in Figure 2.4.



*Figure 2.4: Illustration of X-ray production using an X-ray tube with a tungsten target, electrons flow from the Cathode where the electrons will interact and produce X-rays [24]*

The electron beam is emitted from the cathode through thermionic emission [10]. These electrons are then accelerated across an electric potential from the cathode to the anode. The electron beam illuminates a metallic target which is comprised of a high  $Z$  number material typically tungsten. Electrons will interact with the target in one of two ways, either X-ray fluorescence or Bremsstrahlung [8]. X-ray fluorescence occurs when a high energy electron collides with an electron in an atom's inner shell, resulting in the electron being knocked out of the atom. An electron in a higher orbit will then drop down to the vacated position in the lower orbit and in doing so emit a characteristic X-ray. The energy of this photon depends on the atom. Thus an X-ray emission spectrum for the atom is produced for discrete wavelengths. Bremsstrahlung occurs when an electron is decelerated due to the interaction of the electron with the charged nucleus. The electron path is deflected and decelerated losing energy, this change in energy is emitted as a photon [8]. As the loss

of energy is dependent of the electron to nucleus interaction distance this effect produces a continuous X-ray spectrum ranging from zero energy up to the energy potential which the electrons are accelerated across. This is illustrated Figure 2.5.

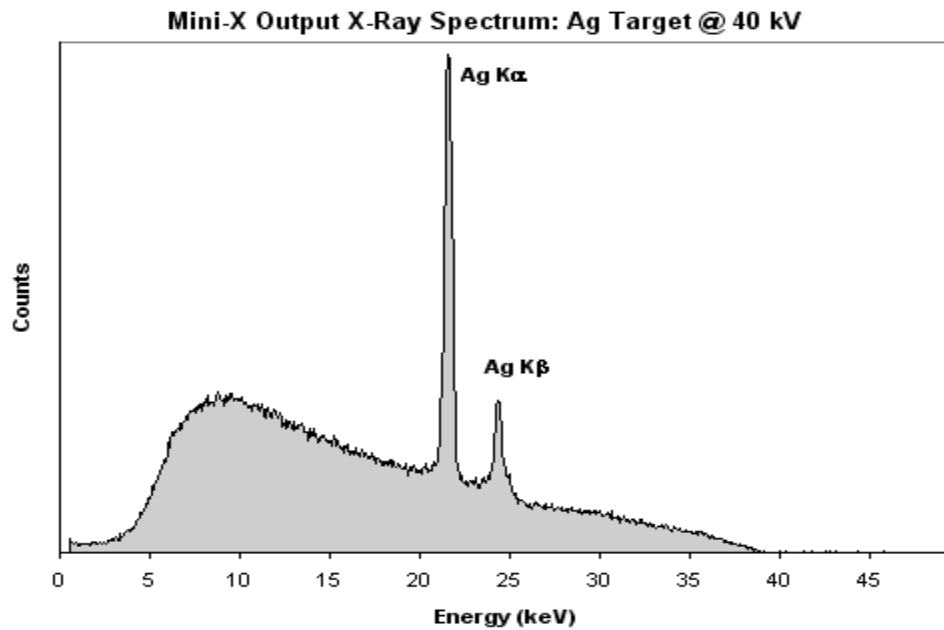


Figure 2.5: X-ray spectrum from a silver target. Ag  $K\alpha$  and Ag  $K\beta$  are characteristic Bremsstrahlung peaks [23]

The two peaks result from X-ray fluorescence and the continuous spectrum results from Bremsstrahlung. X-rays used in medicine are grouped by energies and are split into five different energy ranges; diagnostic, superficial, orthovoltage, supervoltage, and megavoltage. These have the corresponding energies of 20 to 150kV, 50 to 200kV, 200 to 500kV, and 1 to 25MV [8]. Lower energies are used for diagnostic imaging, medium energies for superficial skin treatment, and high energies are used for treating deep tumours. Treating at high energies requires electrons at over 1MeV but accelerating electrons with an electric field breaks down at high energies. Therefore, a linear accelerator is used to accelerate electrons to MeV energies for most therapeutic treatments.

## 2.1.2 Linear accelerators

Linear accelerators use thermionic emission to produce electrons. These electrons are then required to be further accelerated to MeV energies. This is achieved through the use of an accelerating waveguide, either a standing wave or traveling wave accelerating waveguide. The waveguide uses a radiofrequency (RF) electric field to accelerate the electron from the keV range to the MeV range. The electric field is parallel to accelerating waveguide in a traveling wave accelerator and perpendicular in a standing wave accelerator [8].

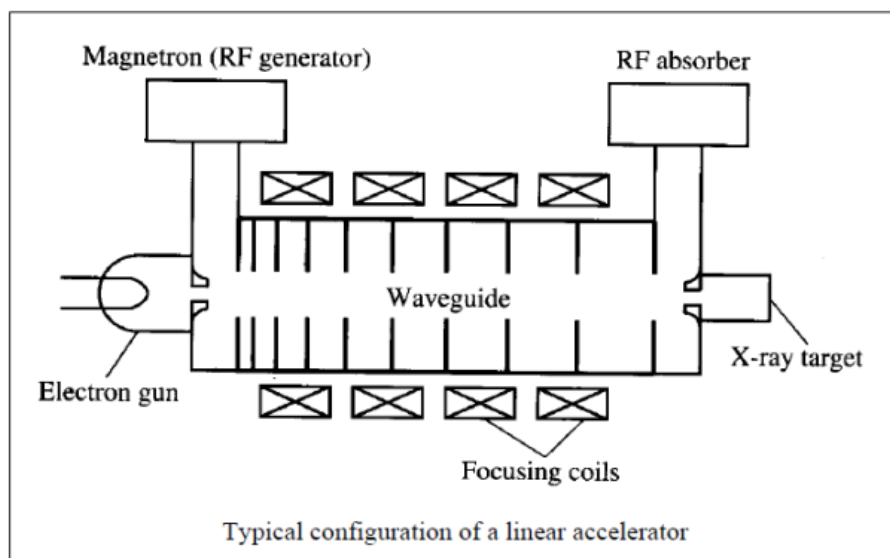
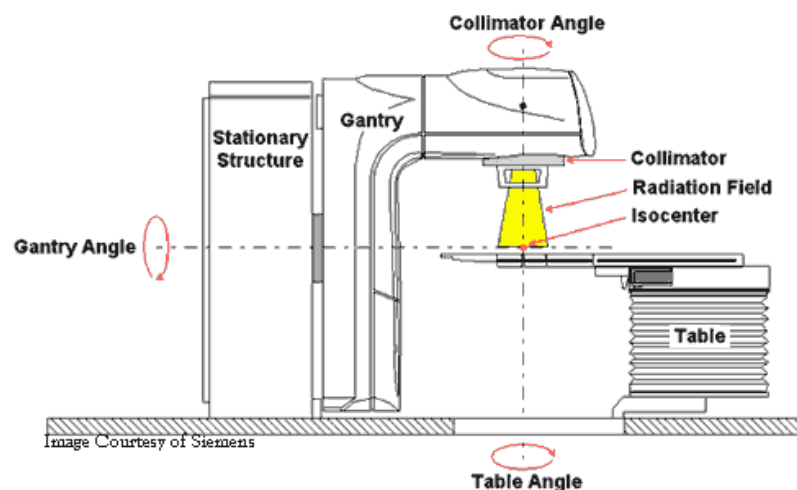


Figure 2.6: Diagram of a traveling wave LINAC displaying core components [46]

This electric field accelerates electrons as they pass through each cavity in the accelerating waveguide. The waveguide in a traveling wave accelerator is shown in Figure 2.6, as the electron speeds increase when traveling along the waveguide from the electron gun the cavities increase in size. For a standing wave accelerator the cavities are a constant size. The RF electric field is generated by the magnetron in Figure 2.6 case, a klystron may be used instead of a magnetron to amplify the radiofrequency electric field. Figure 2.7 is a diagram of a typical LINAC demonstrating the layout of the LINAC and treatment couch.

Patient are positioned on the treatment couch shown as the board attached to the table on the right image. This couch is capable of horizontal, vertical and longitudinal translational motion and rotational motion as shown as table angle. These movements are used to move the patient into the correct position for receiving radiation therapy. The LINAC head and Collimator can be also rotated around the isocentre shown in Figure 2.7 allowing various angles and treatment positions to be produced.



*Figure 2.7: Diagram of a Linear accelerator (LINAC) with a patient couch demonstrating the typical layout and design of linear of treatment machines and the attached couch [45]*

### 2.1.3 Biological interactions

Radiation is used to kill cells through either of two interactions: the ionisation can either occur with DNA or a local water molecule [11]. A collision with a water molecule is more likely, as cells have a greater volume of water compared to the volume of the DNA. A collision between ionising radiation and a cell's DNA causes that particular atom within the DNA to become ionised, which results in a break in the cell's DNA strand [11]. A collision between ionisation radiation and a water molecule causes the water molecule to become ionised, producing free radicals. These radicals can also consequently bond with the DNA strand, resulting in damage to that strand. Damage from a break is easily repaired by using the non-damaged strand as a template. However, if two ionisations occur spatially and temporally close in a DNA strand, the resulting damage to the DNA causes a double strand break [11]. Damage occurring from a double strand break and from free radicals is by and large irreversible, leading to genome rearrangements which ordinarily results in the cell's death.

## 2.2 Cancerous tissue

Cancer as defined by the World Health Organisation (WHO) is "The uncontrolled growth and spread of cell" [12]. This disease can occur in most tissue types and organs in the human body. It is a product of damaged DNA which results in a cell undergoing continuous growth with no means of self-regulation. This will result in development of a tumour. Tumours are classified into two primary categories, benign and malignant. Benign tumours lack the ability to invade surrounding tissue or metastasize. These are of less concern as they can only affect a patient by compressing other tissue.



Malignant tumours however, are capable of invading and destroying surrounding healthy tissue, compressing vital organs and metastasising. Metastasising is a process where part of a tumour breaks off from the primary tumour location and is transported around the body to a new location where a secondary tumour can develop [11]. Primary and secondary malignant tumours destroy healthy tissue through invading nearby tissue. As cancer can regrow from only a few cells it is highly important that the entire cancer is treated to ensure there are no relapses. With this in mind the next section will discuss imaging methods to locate and identify the extent of cancerous tissue.

### 2.2.1 Imaging and detection

Identifying the full extent of a tumour is highly important as all future treatments will be based on this initial information. If the full extent is not identified and therefore not treated, a small section can survive, and if this occurs then the cancer can regrow. Therefore, a range of imaging modalities are used for diagnostic and geometric characterisation of the tumour. Radiotherapy is primarily interested in geometric characterisation of tumours. Patients whom are sent to a Radiation Oncology department have usually already been diagnosed with cancer. Radiation Oncology is interested in geometric characterisation. So a tumour's shape, size and location can be determined and therefore correctly targeted with radiotherapy. The primary imaging modalities used are Computed Tomography (CT), Magnetic Resonance Imaging (MRI). Positron Emission Tomography (PET) and Single Photon emission Computed Tomography (SPECT) are used as functional imaging modalities. All imaging modalities used produce 3D volumetric images. CT and MRI can be used for planning radiotherapy treatments as they produce geometric information and the tumour is visible in these scans. CT are the primary scans used, primarily as the comparative cost of MRI is significantly higher. Depending on tumour type and imaging modalities available, multiple methods are combined to improve identification and localisation of tumours. PET and SPECT can be used in conjunction with CT and/or MRI. PET and SPECT modalities use radioactive isotopes attached to functional pharmaceuticals. These pharmaceuticals are designed to target a tumour's biological makeup. This results in the tumour absorbing a large portion of the radioactive isotopes. Using detectors the location of the isotopes can be determined and therefore the location of the tumour. CT is the primary imaging method for Radiotherapy.

### 2.2.2 Computer Tomography

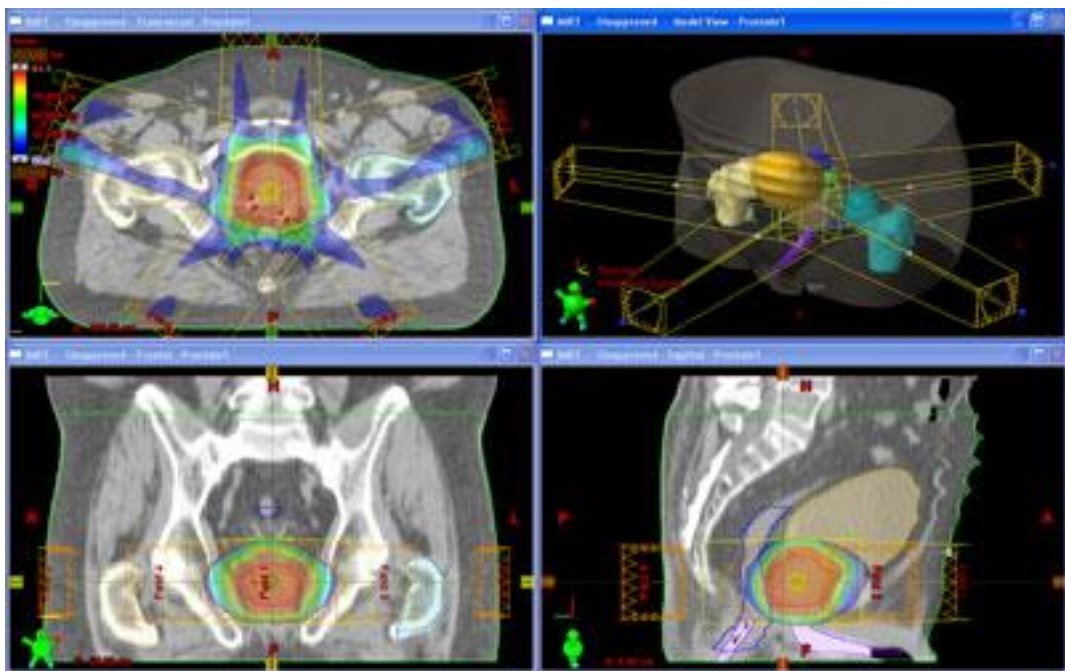
This section discusses CT image production as a minor section of this work was related to surface based CT reconstruction. CT imaging is used to obtain tomographic images of a patient's internal anatomy using X-rays. CT scanners consist of a fanned X-ray source and a detector attached to a ring with a separation between the source and the detector of 180 degrees. A sinogram acquisition of projection images is first produced by rotating the ring around the patient. A Sinogram is an x-y image of transmitted X-rays through the slice against an angle of projection. For an accurate reconstruction to be produced, a rotation of 180 degrees is required. To improve reconstruction quality, high resolution sinograms can be used. Reconstruction of a sinogram to produce a tomographic image uses filtered back projection. This filtered back projection is based on the inverse Radon transform. Production of a full 3D image is achieved by dividing a patient into lateral slices and producing a tomographic image from each slice. In modern CT scanners an iterative process may be used instead. When patients are aligned for treatment on a LINAC, a cone beam CT is may be used. This differs from a normal CT due to a cone of divergent X-rays being used in place of a fanned beam. 4D computer tomography adds a time component to 3D images. This is important for highly mobile tumours located in or around the lungs where the patient's breathing must be taken into consideration. A frame rate is initially selected and since a patient's breathing is in theory a repeating regular near sinusoidal wave, images can be combined into "bins". Reconstruction is done by combining sinogram images into corresponding bins based on either amplitude or phase in the breathing cycle. While this works well if a patient has a regular breathing cycle, distortion can occur if incorrect images are binned together. The breathing cycle is often obtained by placing an infrared reflecting block on a patient's chest, a camera will monitor the

position of this block. The primary issue with this method is it can only track a single point on a patient's chest.

## 2.3 Radiotherapy

Ionisation radiation interaction with biological tissue was discussed in a section 2.1.3. This highlighted that ionising radiation is detrimental to cells. As ionising radiation does not distinguish between healthy and cancerous cells it is harmful to both.

Therefore any treatment needs to decrease the radiation dose to healthy tissue while increasing the dose to the cancerous tumour. This ratio between dose to cancerous tissue and healthy tissue is known as the therapeutic ratio. Increasing the therapeutic ratio will result in superior outcomes for the patient. However, a tumour must receive a specific dose in order for it to be controlled. If this is not realised, then the treatment will not lead to local control. The primary method for improving this is through use of multiple



*Figure 2.8: Screenshots from an intensity Modulated Radiation Therapy Prostate plan. Regions of red demonstrate high dose regions [7]*

beams and beam shaping. Multiple beams can be combined to intersect with the tumour, effectively increasing the therapeutic ratio. This effect is displayed in Figure 2.8 where 5 beams are used to radiate the prostate of a patient for prostate cancer.

The area in regions are colour coded to display isodose contours. With regions in red receiving high dose and regions in blue receiving low dose. As the image shows, the majority of the dose is delivered to the prostate and only a small dose is delivered to healthy tissue. The Beams in Figure 2.8 are also shaped so they conform to the volume of the prostate, this is achieved with beam shaping.

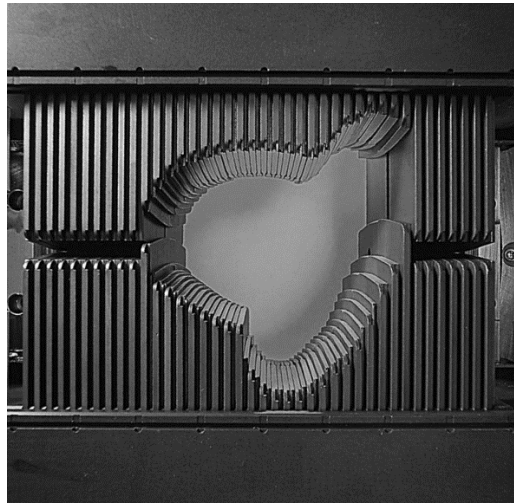
### 2.3.1 Beam shaping

Beam shaping is achieved through the use of collimators. This allows different size fields to be produced, avoiding critical organs and tissue.

Multi-leaf collimators (MLC) are comprised of leaves of metal that can be individually placed in the radiation beam blocking sections of the beam. Collimators are made of thick tungsten leaves which can be placed in the radiation beam effectively blocking sections of the beam. MLCs allow complex two dimensional (2D) shapes to be produced. Figure 2.9 is of a Varian 120-leaf MLC demonstrating how a MLC can be shaped to block sections of the radiation beam. The shape prescribed for beam shaping is based on target volumes determined during treatment planning.

### 2.3.2 Target volumes

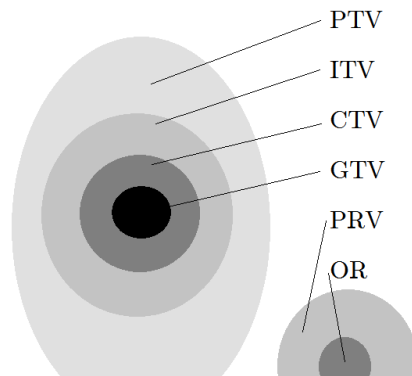
In order for a successful treatment to be planned, target volumes are defined based on imaging which is used to plan the treatment. There are typically six planning volumes



*Figure 2.9: Image of a Varian 120-leaf multi-leaf collimator (MLC), various patterns can be produced by moving the various leaves shown in the image [8]*

determined in radiation oncology according to ICRU 62 [13], the Planning Target Volume (PTV), Internal Target Volume (ITV), Clinical Target Volume (CTV), Gross Tumour Volume (GTV), Planning organ at Risk Volume (PRV) and the Organ at Risk (OR). These are illustrated in Figure 2.10. There are four volumes which are used to describe and plan treatment for a tumour these are the PTV, ITV, CTV and the GTV. The GTV is the location and extent of the tumour which contains the entire tumour visible, palpated or imaged [14]. The primary site, lymph nodes and spread into adjacent soft tissue should all be included in the GTV where applicable [14]. This region may not contain the entire extent of the tumour as the edges are difficult to fully define due to limitations in imaging and identification.





*Figure 2.10: Diagram illustrating the main radiotherapy planning volumes: [46]*

The GTV is surrounded by the CTV this is a volume which takes into account the uncertainty of the GTV and should contain the entire tumour volume. This volume also contains margins for sub-clinical disease. Since this volume is not defined by imaging it is difficult to define, it is defined based on a clinical assessment of future progression and risk of the disease. The ITV expands the CTV by taking into account internal movement for example when a patient is breathing. The ITV is encompassed by the PTV. The PTV volume is based on physical limitations dependent on treatment modality and the linear accelerator. This volume is designed so the prescribed dose which is to be delivered to the CTV can be achieved. This is resulting from the reality that radiation fields outputted from a linear accelerator are not completely uniform, have associated uncertainties, associated setup uncertainties and are impacted by tissue which the radiation beam passes through. There are two volumes which describe organs at risk, the OR and PRV. Organs at risk are organs nearby the tumour which cannot receive a high dose of radiation therefore must be avoided by radiation beams. The OR volume contains the organ at risk similar to the GTV. The PRV contain the OR. It allows for organ movement and any inaccuracies with defining the entire organ. This volume is similar to the ITV. Using these six volumes, treatment planning can ensure at risk organs are spared and the entire tumour receives the prescribed dose.

## 2.4 Radiotherapy Modalities

This section discusses the advancement of modern modalities for radiation therapy progressing from 3D Conformal Radiotherapy (3DCRT) to Intensity Modulated Radiotherapy (IMRT) and finally onto Stereotactic Body Radiotherapy (SBRT).

The advantages of each of these methods will be discussed in this section. As technology improves, the PTV decreases so to match the CTV. 3DCRT is the oldest of the modalities discussed in this thesis however it is still in widespread use throughout the world. 3DCRT uses multiple beams for treatment. Each beam is shaped from the beam's eye view (BEV), so that the beam fits the 2D projection of the target. The combination of multiple 2D shaped beams will result in the tumour receiving a full three dimensional (3D) dose. Each beam delivers a percentage of the total dose delivered to the tumour. The intensity of each beam is uniformly varied, or weighted, based on the volume of critical structures the beam passes through.

IMRT is an improved methodology based on 3DCRT. IMRT further shapes the beams and the intensity non-uniformly of each beam to ensure a uniform dose is delivered to the tumour. This also decreases the dose delivered to healthy tissue. This is primarily achieved with use of inverse planning. A total dose volume is prescribed to a tumour target area. Constraints are based on dose limits to critical structures. With this information computer algorithms can calculate the radiation beam intensity and collimation shape required to deliver the specified dose to the tumour and avoid critical structures. This results in patients being treated with multiple smaller segments.

SBRT is a relatively new methodology which uses high dose beams. SBRT uses smaller margins compared to centimetre margins used by 3DCRT and IMRT [15]. SBRT offers an increase to the therapeutic ratio and allows very small radiation fields to be produced. However, due to the very high dose gradients, great care is required as even a very small misalignment will result in a large dose delivered to healthy tissue, and the tumour will receive too small a dose. While the physical advantages are well understood, the Radiobiological understanding is “Poorly understood” [15] as this is an emerging modality.

Image Guided Radiotherapy (IGRT) is used in conjunction with other treatment methods to improve treatment results. IGRT uses imaging to ensure the patients tumour and internal organs are positioned correctly before the radiation beam is turned on. Such that the current position the tumour is located coincides with the planned position from the treatment plan. If this does not occur, healthy tissue will receive a high dose resulting in a low therapeutic ratio and negative treatment outcomes.

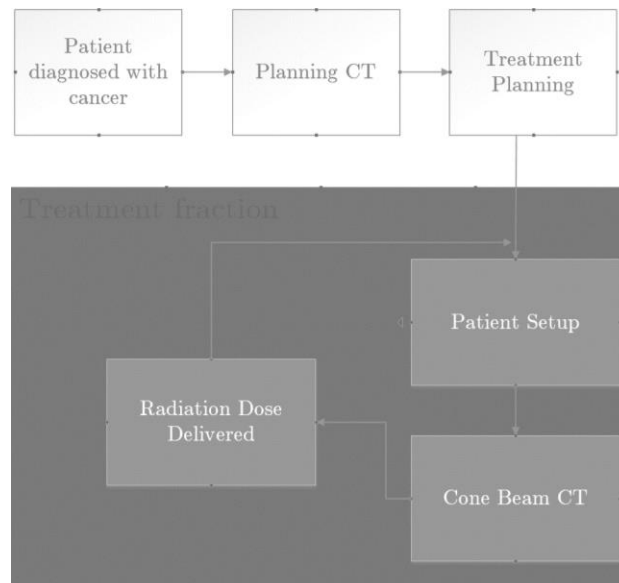
As internal tumour's position do not necessarily move with the external skin methods to observe the internal positions are required these methods are usually on board Cone-beam CT (CBCT) built into the LINAC or X-ray imaging. Images obtained are then compared to images from treatment planning images. Patients can be aligned with 2D, 3D or four dimensional (4D) imaging or a combination of these. Use of higher dimensional imaging allows the tumours position to be more accurately localised. 4D imaging produces a 3D image which can vary over time. This can be important for tumours located in highly mobile sites, for example, lung tumours. These tumours are very mobile due to the patient breathing throughout treatment. Typically after a patient is setup in the treatment couch in what is believed to be the correct position for treatment, a cone beam CT scan is performed. This scan is then compared to the treatment plan, if the plans coincide then treatment can commence, however if they do not then the patient will need to be moved this can be achieved by either moving the treatment couch or by moving the patient directly. After the patient is moved to the new position the patient will be reimaged to ensure their position is now correct. This method using cone beam CT is a highly accurate method to ensure a patient is positioned in the correct position before treatment, however it will only take into account the position during the scan and is not effective if the patient moves at a later stage.

As treatment methods advance from 3DCRT to SBRT margins on treatment volumes are tightened, this is achieved through improvements in LINAC design and improved imaging. However as a result of this, patient motion has an increasing impact on treatments increasing the need for patient motion management which will be discussed in the next section.

## 2.5 Patient motion management

This section starts with an overview of the clinical workflow present in a radiation oncology department, as this directly relates to the patient positioning issues and patient motion problems. The common workflow for a patient going through a Radiation Oncology clinic is shown in Figure 2.11. The process starts with a patient being diagnosed with a cancerous tissue and then recommended to receive Radiotherapy for their cancer treatment. The patient will then require a planning Computer Tomography (CT). This initial planning CT forms the reference image from which all treatment planning is performed. Therefore, when receiving radiotherapy at a later stage, the patient is set up so their position coincides as well as possible with this initial position.

This is done by tattooing external marks on a patient's skin during the planning CT. The CT scan is performed to identify the extent of the tumour and is used to plan the treatment.



*Figure 2.11: Illustration of patient workflow in a Radiation Oncology department, Treatment fractions shown in the large grey box are repeated daily over a few weeks*

The planning phase is also used to identify critical healthy regions that have low dose tolerances. These critical regions and the tumour are then contoured on the CT images. This is then used to plan the treatment. The radiation treatment is usually split into multiple 2 Gy fractions which are delivered each treatment session over the course of a treatment. The total dose is dependent on the type, size and location of the tumour. Fractions are delivered daily over the course of a few weeks. For each fraction, a patient is placed on the treatment table in the Linear accelerator (LINAC) bunker. The patient is then aligned to room lasers using the marks tattooed during the planning CT scan. A cone beam CT scan is then performed to ensure the patient's 3D anatomy aligns to the planning CT. Once the patient is in a position that best matches the initial position, the treatment fraction is delivered to the patient.

### 2.5.1 Patient Positioning

Prior to treatment of cancer using external beam Radiotherapy, a patient is positioned on a treatment couch in the supine position. This is initially a 3 point alignment with room lasers. A more precise alignment is performed with a CBCT scan. This scan produces a three dimensional (3D) image of the patient. This ensures the patient is in the correct position at the start of treatment. After this alignment is complete, the patient receives the prescribed dose.

However, after the initial setup, there are currently few systems that monitor the patient's position over the entire course of treatment. This can lead to a patient moving from the correct position over the course of a treatment fraction, and this may result in radiation which is prescribed to the tumour potentially being delivered to healthy tissue. Should this happen it will have a negative impact on the patient's treatment.

## 2.5.2 Patient motion management

The methods of motion management differ depending on the tumour location. For head and neck cases, a mask which has been moulded to a patient's head locks the patient's head in place, preventing or at least minimising any movement. This method while uncomfortable for the patient allows for easily re-creatable positioning. Vacuum cushions can also be used to correctly position the patient's entire body. These cushions become rigid when a vacuum is applied, allowing them to be placed underneath a patient and be moulded to fit the patient. These cushions rest in setup error of between 1–9mm depending on tumour site [16]. However, they offer a simple yet comfortable alignment method. The primary issue regarding patient motion management is patient breathing. As it is not possible to completely stop the patient's breathing throughout the treatment any tumours located in or close to the lungs will be highly mobile. There are currently three primary methods for addressing this, abdominal compression, breath hold or gating. Abdominal compression limit patient breathing, by applying pressure on the patient's abdomen. Breath holds stop the patients breathing for a short period of time, either through active systems where the patient's air pathways are blocked or passive systems where the patient is asked to hold their breath for a period of time. Gating monitors the patients breathing throughout their breathing cycle and turns the radiation beam on and off based on when the patient's chest is in the correct position. These three methods can be used in conjunction. An Example is the Active breathing control (ABC) system which consists of a device which is placed in the patient's mouth that the therapist can turn on to block airflow placing them in a breath hold. This is used in conjunction with gating, where the beam is off when the patient is not in a breath hold.



A study on the reproducibility on organ position using the ABC system for liver therapy by Laura A Dawson et al [17] demonstrated that the intra fractional reproducibility of the diagram position relative to the skeleton for 90% of patients in the study was less than 5.4mm.

### 2.5.3 Surface tracking

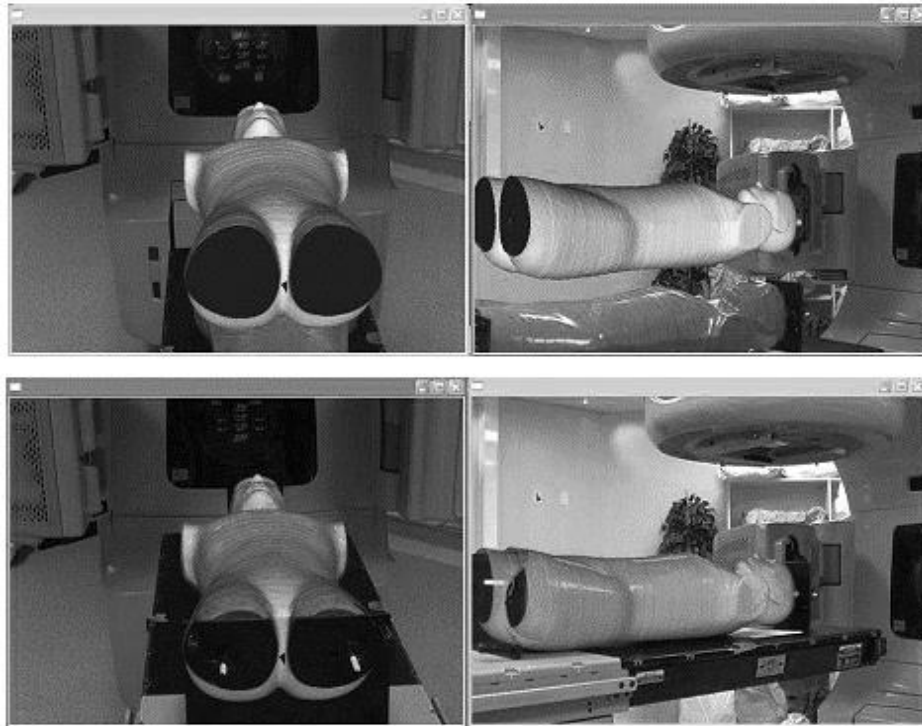
Currently there are a two commercial surface tracking systems on the market, capable of tracking external surfaces of a patient throughout treatment. However, these systems are not commonplace. The two primary patient positioning and monitoring systems are the Catalyst system produced by C-RAD (Uppsala, Sweden) and the AlignRT system produced by VisionRT (London, United Kingdom). These systems use depth sensing cameras to acquire 3D surfaces of a patient in real-time and compare them to the initial position obtained when the CBCT was performed to determine the relative patient movement from the initial position. Details on how depth cameras operate is introduced in chapter 3.

## 2.6 Previous work

To assist with aligning the patient to the planning CT and monitoring the patient throughout treatment Talbot [18] developed an Augmented Reality system. This section discusses this solution and why a different approach was required.

The system "A Patient Position Guidance System in Radiotherapy Using Augmented Reality" developed by Talbot [18] uses augmented reality to visually align the external contour of the patient. This system was developed at the University of Canterbury as a prelude to the current project. In Talbot's project the augmented reality real world components are acquired by a real time video stream input from a camera. In Talbot's projects case a camera is placed in a LINAC bunker viewing a patient being setup on a treatment couch. A 3D model of the patient's surface obtained from the CT data from SIM is augmented into the scene, demonstrating the correct position relative to the radiation Isocentre. This is demonstrated in Figure 2.12. The patient should be aligned to this model so they are in turn aligned correctly to the radiation isocentre.

The 3D model is positioned in the correct position through the use of tracking markers. The AR system allows patients to be visually aligned to the correct position from the cameras point of view (POV). With the use of two camera viewpoints a patient's position can be aligned in 3D to the correct pose. This system used two web cameras to obtain a qualitative 3D alignment. However, a lack of quantitative position information has limited the system, and as the alignment is a relative alignment the only information available is whether a patient is misaligned or has an incorrect pose or is incorrectly setup. There is no numerical information on how accurate the alignment is or the relative movements required to correct any position offsets.



*Figure 2.12: Screenshot of Talbot's system, demonstrating the process of aligning the real object to the 3D surface, the top two images show the initial unaligned setup and the bottom two images show the correct alignment [13]*

Numerical information is important in situation when a patient has moved from their initial position, the direction and magnitude of the movement detected from a patient monitoring system would be useful when correcting this movement. Therefore a depth camera setup is required. Use of a single 3D depth camera also removes the complication of having a dual camera system and removes the need for tracking markers to be used.

# Chapter 3: Principles of Depth Cameras

---

This chapter introduces the various types of depth cameras required to provide 3D analytical information about patient position. Previous work [18] has demonstrated the need for analytical information which depth cameras can provide. An overview of the current methods of depth detection is provided and information on the two cameras used in the following sections of the thesis.

## 3.1 Depth Cameras

Currently there are three primary methods for producing a 3D surface scan of an object: a contact scanner, where the scanner physically measures the depth at each point, a non-contact passive scanner, which relies on detecting ambient radiation, or a non-contact active scanner which actively produces radiation to scan the room and detect with a detector [19]. Of these methods, the contact scanner is unsuitable in a medical environment and would be uncomfortable with a patient and the accuracy of images produced from the passive scanner is limited. The dependence on external lighting also limits their use for this application. Therefore non-contact active scanners will be used for this project due to their suitability in a medical environment, the accuracy of the images produced and extra or external lighting is not required. The following sections will introduce the different methods and types of non-contact active scanners.

### 3.1.1 Time of Flight

Time of flight depth cameras use the known speed of light to measure distance [20]. They send out a light pulse from a laser and record the time from when the light signal is sent till the time the reflected signal is detected by the sensor. Since the speed of light is constant as the medium does not change, the total distance travelled by the photon can be calculated and half this distance is the distance between the camera and the object. This is described in Equation 2.1 where  $D$  is distance from the camera to the observed object,  $t$  the total time travelled by the pulse of light and  $c$  the speed of light (approximately  $3 \times 10^8 \text{ ms}^{-1}$ ) [20]

$$D = \frac{1}{2} ct \quad (2.1)$$

As a consequence of the large value of the speed of light, in order to detect an object at a distance of three metres away, the camera must be able to record the time of flight down to 10ns. This method will produce a single distance point measurement which is used in laser range finders. To produce a full 3D scan of an object, the laser can be swept in the x-y directions, producing a full 3D map. There are two time of flight methods that sweep the laser over a 3D surface. Either with the use of rotating mirrors to change the angle of the laser beam or by rotating the laser itself. Rotating mirrors are commonly used due to simplicity [21]. Mirrors are significantly lighter than the whole laser and thusly can be rotated at a higher rate, decreasing the total scan length. Rotating a mirror at a high speed enables a rapid scan rate of a hundred thousand points per second [21]. This method will generate highly accurate measurements over a large static scene. However, as each point is sampled individually, the resulting total scan time required to produce a 3D image is very large, giving this method a low associated frame rate [21].

Instead of scanning over the whole scene one point at a time, a whole 2D matrix of depth points can be collected at once. This is known as a Time of Flight Camera TOF. This method is significantly faster and has a scan rate in the order of millions of points per second.

### 3.1.2 Time of flight camera

Time of flight cameras work using the time of flight principle as discussed in the previous section. However, they are also able to collect a 2D grid of depth information and therefore achieve a high frame rate. Typical time of flight cameras have a horizontal and vertical resolution of 320x240 [22]. For example, the commercially available TOF camera by Microsoft™ called the Kinect v2 is able to collect 6.5 million points per second [23], significantly exceeding the data collection rate of the single point method by a factor of at least 10. There are two main technologies which time of flight cameras use, RF modulation and Range gating [20]. RF modulated cameras use a radio frequency to modulate the outgoing light beam, the phase of the signal detected is compared to the internal phase, resulting in a phase shift for each pixel.

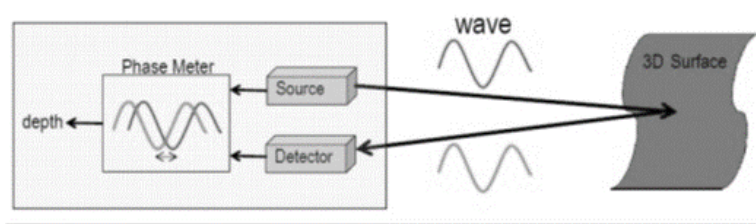


Figure 3.1: Diagram of a RF modulated TOF camera, showing the outgoing wave from the source and the resulting wave reflected off the 3D surface. [42]

This is illustrated in Figure 3.1 where a source outputs a wave to the 3D surface and the phase meter. The detector receives the reflected wave which can be compared to the original source wave providing a depth measurement.

Range gated cameras use a shutter which is opened and closed at the same rate as

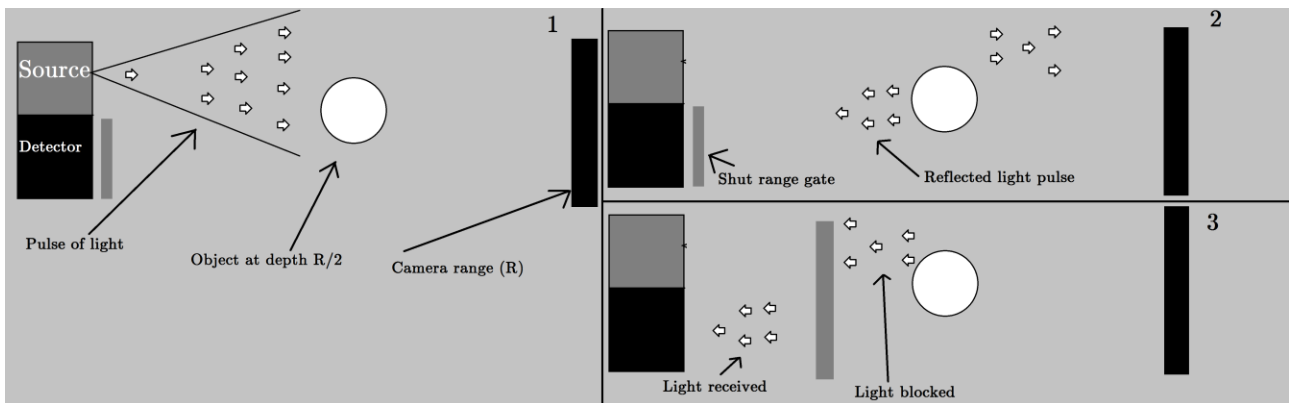


Figure 3.2: Diagram of range gated time of flight camera, illustrating the pulse of light moving from frame 1 to 3, reflecting off the white sphere and being detected by the detector.

pulses of light are sent out, resulting in a proportion of light being blocked. By comparing the difference between light received and light blocked, the distance can be calculated [20].

Figure 3.2 illustrates the principle operation of a range gated time of flight camera. The pulse of light is emitted with a known intensity of light in this diagram ten packets of light. These packets of light will be reflected off any objects within the camera range. The gate will then let a portion of light through and block any remaining light. The following equation (2.1) can be used to calculate the distance between the object and the camera.

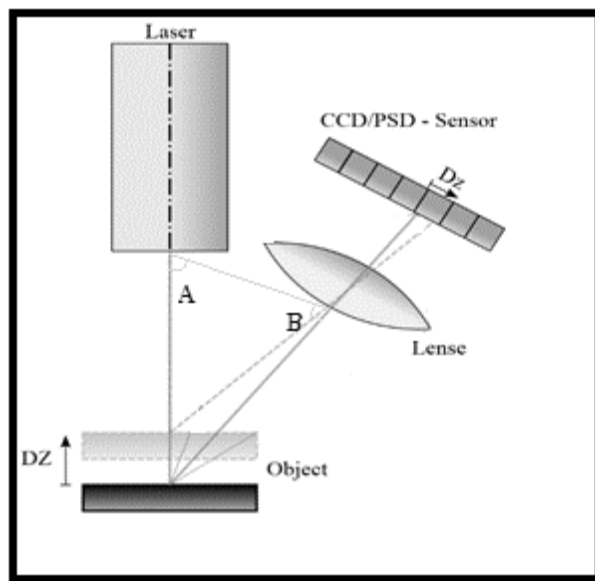
$$Z = R \frac{S_2 - S_1}{2(S_1 - S_2)} + \frac{R}{2} \quad (2.1)$$

Where  $Z$  is the distance,  $R$  the Cameras range,  $S_1$  light received and  $S_2$  light blocked. As  $S_2$  and  $S_1$  are equal the distance equals  $R/2$ . The Kinect v2 uses range gating technology.



### 3.1.3 Triangulation

Triangulation is another method of non-contact active scanning. This uses a radiation emitter and detector system similar to the TOF camera system. However, instead of measuring time taken for the photon to travel, the camera and projector are separated by a known distance. This forms a triangle where a change in depth will cause a shift in the detected pixel on the cameras CCD/CMOS sensor. This is illustrated in Figure 3.3, using the triangle in image the distance  $DZ$  can be found. Given that the distance between the laser and sensor is known, the angle cornered on the laser angle  $A$  is fixed and the angle cornered on the sensor angle  $B$  can be determined by which pixel on the camera detected the laser signal, therefore the distance  $DZ$  can be found. This process produces



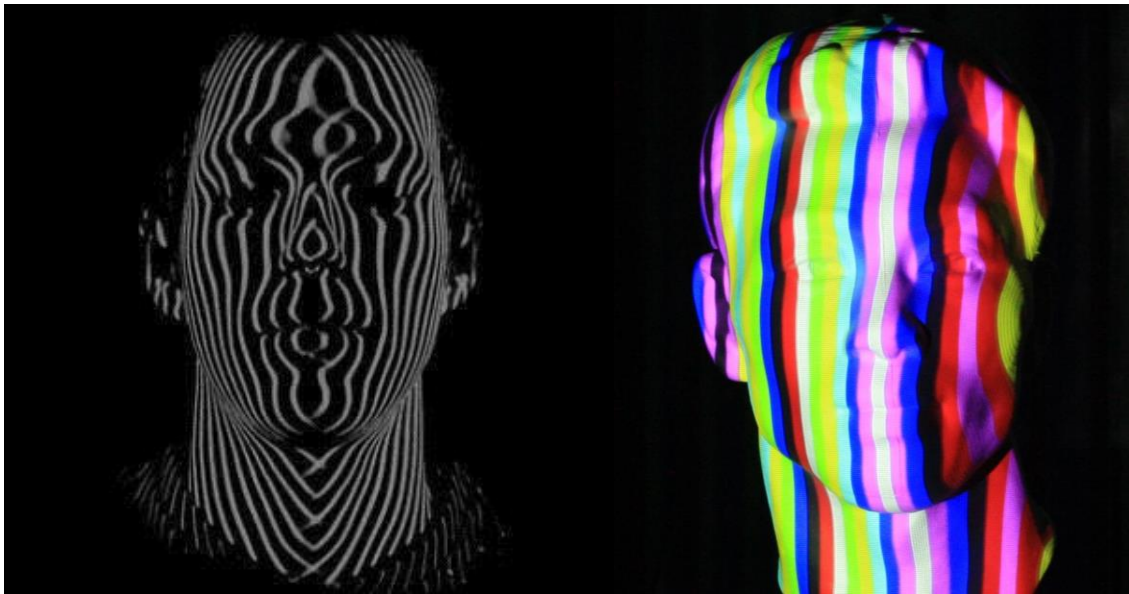
*Figure 3.3: Illustration of a triangulation camera, using a laser and a sensor with a known triangle to determine the distance [33]*

a single depth point.

There are two techniques which produce a full scan of an object with triangulation: temporal encoding or spatial encoding [20]. Temporal encoding has a time component dependence when producing a scan, so individual frames are combined to produce a complete scan. There are two methods of temporal encoding, point scanning and line scanning. Point scanning requires the laser dot to be swept over the whole object in the x and y planes to produce a full scan. This is done in a similar manner as point scanning by the time of flight camera. Line scanning is where instead of scanning a single laser point at a time, a whole laser strip can be scanned at once using a 2D CCD/PSD sensor to improve acquisition time. This only requires the laser line to be swept over the object in the x direction for a whole scan to be produced. Temporal encoding can have the benefit of producing highly accurate scans, but has the detraction of total frame rate. Spatial encoding is a method that uses light encoding. Light encoding can use parallel lines projected on to an object. Each line needs to be uniquely identified either by counting

the lines using pattern recognition, varying the lines wavelength, varying the stripe thickness or another form of encoding.

This is demonstrated in Figure 3.4 consisting of parallel white lines and parallel coloured lines incident on a human face. These lines allow the camera to determine the depth along each line and produce a depth measurement. Some cameras vary from this method.

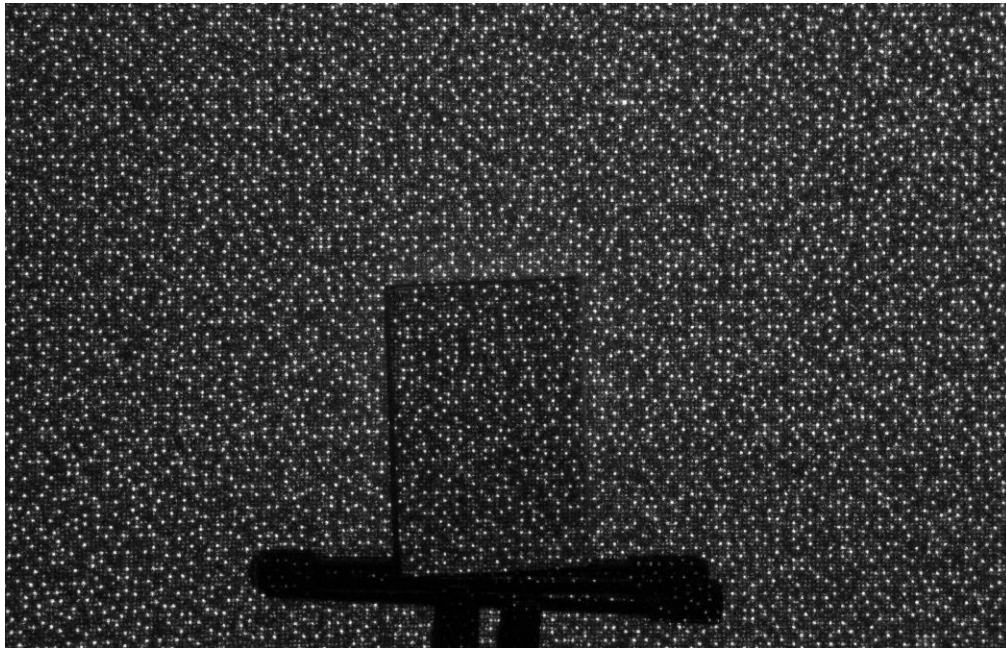


*Figure 3.4: Illustration of parallel line encoding using white lines (left) and coloured lines (right) on a person's face [34] [35]*

Instead of only encoding in the x direction and using parallel lines, a full 2D encoded projection is done where each area is unique. With both encoding methods the position of all the lines and points are known from either spatial or temporal encoding, therefore a full 3D map of an object can be produced.

The Kinect for Xbox™ 360 (Microsoft, Redmond, USA) known as the Kinect v1 uses a 2D encode consisting of a speckle dot pattern [20]. This pattern appears to be random in nature, however some repetition has been observed, showing it is a known unique pattern. This pattern is displayed in Figure 3.5.





*Figure 3.5: Image of the Kinect for Xbox 360 structured light pattern showing a box on a table positioned in front of a wall [36]*

## 3.2 Kinect for Xbox™ 360

On the 4<sup>th</sup> of November 2010 Microsoft<sup>TM</sup> released the Kinect for Xbox. This device was designed to be a motion sensor input device which is capable of capturing user movements while playing games. The Kinect for Xbox 360, also known as the Kinect v1, is based on the non-contact active scanning camera using triangulation. After the release of the Kinect v1, Adafruit Industries promoted development of an Open source driver for the Kinect v1 [24]. On the 10<sup>th</sup> of November 2010 an open source driver was released by Hector Martin allowing for the first development on the Kinect v1 [24]. Prime Sense the company which developed the sensor for Microsoft<sup>TM</sup>, released an open source driver and motion tracking middleware later in December 2010. Official drivers and a SDK for windows was released by Microsoft<sup>TM</sup> on the 16<sup>th</sup> of June 2011. The Kinect v1 consists of a 640x480 resolution Red Blue Green (RGB) Camera operating at 30Hz, a 640x480 resolution Depth camera operating at 30 Hz, a Depth projector, a Multi-array microphone and a motorised tilt. The Depth camera and projector operate in the infra-red spectrum, so are not visible to human eyes at 830nm. The Kinect v1 has a default operating range of 800mm to 4000mm or a near mode operating range of 500mm to 3000mm.

This range is ideal in the clinical environment as the camera will likely be positioned on the ceiling at the foot of the treatment couch in a LINAC bunker. Figure 3.6 displays the Kinect v1 sensor that was used in this project on the top.



*Figure 3.6: Image of the Kinect v1 (top) and Kinect v2 (bottom) sensors [37]*

### 3.3 Kinect for Windows v2

The Xbox One was released on the 22<sup>nd</sup> of November 2013. This new console came bundled with a new Kinect known as the Kinect for Xbox One. The sensor is shown in Figure 3.6 on the bottom. Microsoft™ is currently in development of the SDK and drivers for the Kinect for Windows v2 sensor based on the Kinect for Xbox one. At present Microsoft TM has not released to the general public a Kinect for Windows v2 sensor, also known as the Kinect v2, which will allow use of this sensor with a computer. However, currently Microsoft TM is running the Kinect for Windows v2 alpha program. This program allows early access to the Kinect v2 sensor, SDK, driver and support from the development team however, since this sensor and software is only in an alpha stage of development, all results are preliminary and are subject to change.



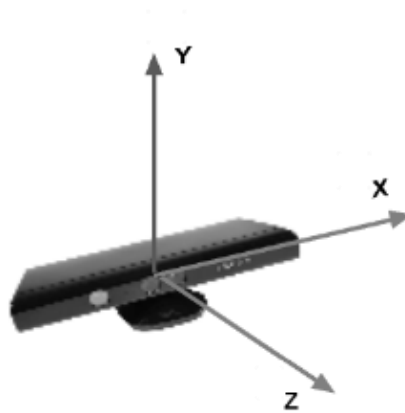
This project is currently part of the alpha program so has had access to the Kinect v2 sensor. The new Kinect v2 depth sensor uses time of flight technology and has improved depth resolution and accuracy. The new depth camera has a resolution of 512x424 pixels, and the improved RGB camera has a resolution of 1920x1080(1080p). The sensor has a similar microphone array and has the tilt motor removed as the field of view of this new sensor has been increased. The time of flight technology used for this sensor is assumed to be a range gating camera though that remains an unknown at time of writing as Microsoft™ has not yet released the full specifications of this sensor. Therefore all information concerning the internal processing and technology may be inaccurate. It is believed that the internal frequency of this sensor is 300Hz, however only a 30Hz output frequency is provided through the SDK. The Kinect for Xbox One which is the same sensor, is capable of running at 60Hz, so it is assumed that the Kinect for Windows v2 will be able to achieve this at some time in the future.

This chapter introduced the different technology used in various depth cameras introducing the two sensors which this project will be utilising, the Kinect v1 and the Kinect v2. The Kinect v1 and v2 were chosen for this project due to their low cost and high resolution compared to other depth cameras [22]. The use of depth cameras allow 3D information about a scene to be obtained however objects need to be detected, distinguished and tracked. The next chapter introduces the need for a tracking algorithm in order to obtain the correct position of a 3D object.

# Chapter 4: Tracking Methods

---

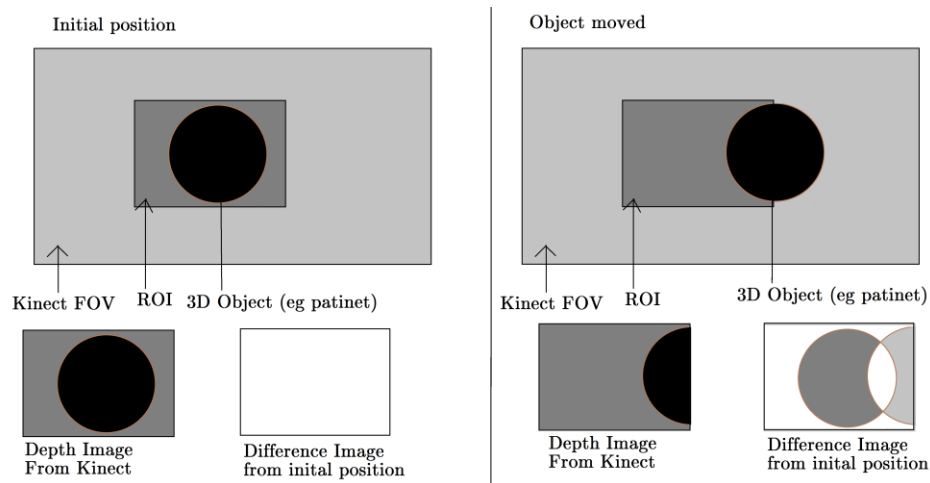
This chapter introduces the need for tracking software in the system in order to detect x



*Figure 4.1: Diagram of the Kinect coordinate system used by both the Kinect v1 and v2 [43]*

direction and y direction motion, and will discuss the various methods of tracking which were investigated for this project. Figure 4.1 illustrates the Kinect v1 and v2 coordinate system used.

Raw data obtained from the Kinect is in the format of either an x-y depth image or an



*Figure 4.2: Illustration of object movement in depth images. A black object is observed at an initial position which moves to a different position at a later time. As the ROI does not follow the object the real position cannot be found.*

x-y RGB image, where each pixel corresponds to a measurement of depth from the initial position. A mask is placed over the depth image to constrain the image to a small region of interest (ROI) which will be tracked. A depth mask is also applied, removing all points which are outside a set range. This range is based on a volume which the patient should be located and is required to reduce noise in the system from incorrect depth pixels. Tracking an object's movements in the sensor's z direction is trivial as the mean difference between the depth images is the movement in the z direction. However, a movement in the x or y directions will cause the object to partially or fully move outside the mask. Figure 4.2 illustrates this problem. Showing an object moving from its initial position to a new position outside the ROI, from the sensors point of view there was no movement in the x direction only a change of depth values.

Therefore, a 2-D tracking solution is required to determine the object's position in the whole sensor image. This chapter introduces various tracking methods investigated during this project. An initial literature review was performed and a range of tracking methods were investigated. These methods were chosen to represent the range of different tracking methods available. Tracking methods were also selected based on ease of implementation. A selection of the most appropriate methods are discussed in the following sections.

## 4.1 Mean Shift

The Mean Shift algorithm is a well-known and established method of tracking an object and was, first introduced in 1975 by Fukunaka and Hostetler [25]. The Mean shift algorithm is an iterative mode seeking algorithm. This method calculates a nonparametric density gradient through the use of a generalised kernel. A nonparametric statistics are based on no assumptions about the underlying data in terms of size or probability distributions therefore this algorithm can be applied to a wide range of data sets including expanding sets. Density gradients is a measure of spatial variation in density over a given area, this can be used to determine the position of objects in an image. Generalised kernels is a weighting function used to estimate the density gradient rather than implicitly determining it.

The mean shift algorithm assumes that all feature spaces are empirical probability density functions, by assuming inputted data points of the feature space are samples of the underlying probability density function. By making this assumption all data inputted is assumed to be a subsample of the entire scene, in the case for a Kinect sensor this is illustrated as the finite resolution available. Dense clusters or regions of data points correspond to local maxima of the probability density function. These maximum are used to determine the position of a tracked object. A window is defined around each data point and the mean is calculated. The windows centre is then shifted to the calculated mean position. This process is then repeated until a convergence is found.

A Rectangular kernel was used for the mean shift algorithm in this project. Let  $K$  be a flat kernel in  $X$  then

$$K(x) = \begin{cases} 1, & a \leq x \leq b \\ 0, & \text{else} \end{cases}, \quad (3.1)$$

This is the mathematical representation of a ROI window, removing all data outside the ROI. The Parzen–Window Density Estimation technique is used to calculate the estimated density of a random variable the equation related to this is given in (2.2). For the kernel  $K$  with the bandwidth parameter  $h$  the Kernel density estimator for  $d$ -dimensional points is

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right). \quad (3.2)$$

The mean shift is based on the gradient ascent of the density contour where

$$x_1 = x_0 + \eta f'(x_0). \quad (3.3)$$

Where  $\eta$  is the step size,  $x_1$  is the new position,  $x_0$  is the original position and  $f'$  is the gradient. Equation 3.3 can be applied to 3.2 giving

$$\vec{x} = \frac{\sum_{i=1}^n K'(\frac{x-x_i}{h}) \vec{x}_i}{\sum_{i=1}^n K'(\frac{x-x_i}{h})}. \quad (3.3)$$

This is applied to each data point until convergence on a local maxima is achieved. The resulting stationary points are the local maxima of the probability density function. By assuming that

$$g(x) = -K'(x), \quad (3.4)$$

The resulting equation is

$$m(x) = \frac{\sum_{i=1}^n g(\frac{x-x_i}{h}) x_i}{\sum_{i=1}^n g(\frac{x-x_i}{h})} - x. \quad (3.5)$$

Where  $m(x)$  is the mean shift. Therefore in summary for each data point  $x_i$  the mean shift vector  $m(x_i^t)$  is calculated. The Parzen-Window Density Estimation is then shifted by  $m(x_i^t)$ . This process is repeated until a convergence is found. The mean shift processing time is based on  $Tn^2$  where  $n$  is the number of points and  $T$  is the total number of iterations. "The mean shift algorithm was never intended to be used as a tracking algorithm, but it is quite effective in this role." [26]

## 4.2 Camshift

Continuously Adaptive Mean Shift (Camshift) [26] improves on the mean shift tracking method by including scale tracking and rotation tracking. This improvement was developed by Gary Bradski while working for the Intel Corporation [26].

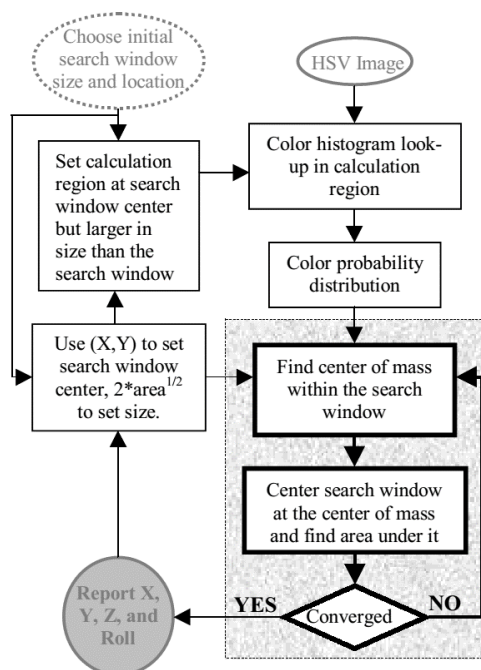
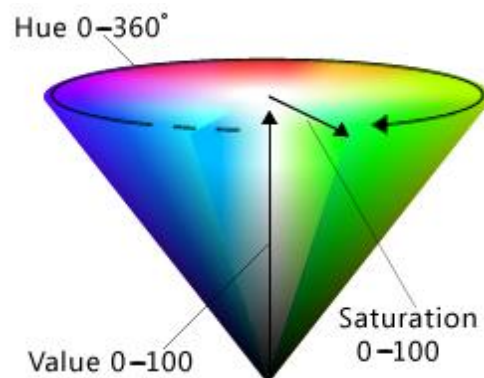


Figure 4.3: Flow diagram of Camshift Tracking method showing the Meanshift method in the grey box

This implementation has been developed for computer vision in particular. Figure 4.3 illustrates the Camshift method. This method encompasses the mean shift algorithm which is shown in the grey box. The Camshift method runs on each video frame. Hue Saturation Value (HSV) colour space is used for colour images instead of the Red Green Blue (RGB) colour space. For this method, HSV is a cone colour space instead of the cubic RGB space, so a conversion is required between them. The HSV colour cone is illustrated



*Figure 4.4: Illustration of the HSV colour space, colours are separated by hue, intensity of the colour is represented by saturation and brightness is represented by value.*

in Figure 4.4. This also shown the dependence on lighting as lighting decreases the value decreases reducing differences between Hues.



An initial window is selected, this window contains objects that the user wishes to track. From this initial window, a search area is set. This bounds how far an object can move between frames. If this outer boundary is not large enough, then a fast moving object could be missed by the tracking algorithm, increasing the search window's size however, this has the drawback of increasing processing time. A 1-D histogram is produced from the Hue channel in the HSV space image. The primary reason HSV colour space is used is, all humans except for albinos have similar Hue values regardless of skin colour; skin colour will only impact on the colour saturation – which is not used. This feature was considered as a key factor when the CAMSHFIT method was selected for analysis in this project. HSV does have the limitation of in a low light situation the diameter of the HSV cone decreases, therefore the differences between hue values will decrease resulting in difficulty in distinguishing between different hue values. After the initial histogram is produced, this is stored and used as a lookup table for all future frames when tracking. Machine learning can be used to improve tracking by combining histograms for each person tracked. This however, was not implemented in this project.

Images are converted to probability values by using the initial histogram as a look up table. The probability values will range from 0.0 being false and 1.0 being true for a match between the pixel and the histogram. This will then form a probability image. The probability image is fed into the mean shift algorithm to determine convergence between the centre of mass and the centre of the window. The size of the new tracking window is then determined by the scale factor  $s$

$$s = \frac{\sqrt{area_1}}{\sqrt{area_0}}, \quad (3.6)$$

Where  $area_0$  is the old area of the tracking window and  $area_1$  is the new area determined from the new placement from the mean shift algorithm. This process is then repeated for any new HSV frames. To determine the rotation of the image around the z or sensor axis the raw moment is used. Which is defined as

$$M_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy, \quad (3.7)$$

As the tracking window has a finite size the formula can be reduced to

$$M_{pq} = \sum_x \sum_y x^p y^q f(x, y). \quad (3.8)$$

The orientation of the semi-major axis is then given by

$$\theta = \frac{\arctan\left(\frac{2\left(\frac{M_{11}}{M_{00}} - x_c y_c\right)}{\left(\frac{M_{20}}{M_{00}} - x_c^2\right) - \left(\frac{M_{02}}{M_{00}} - y_c^2\right)}\right)}{2}, \quad (3.9)$$

Where  $x_c y_c$  is the centre of the window. The semi-major and semi-minor axes can be calculated with the Camshift algorithm. Let

$$a = \left(\frac{M_{20}}{M_{00}} - x_c^2\right), b = 2\left(\frac{M_{11}}{M_{00}} - x_c y_c\right), c = \left(\frac{M_{02}}{M_{00}} - y_c^2\right), \quad (3.10)$$

Then the semi-major  $l$  and semi-minor lengths are

$$l = \sqrt{\frac{(a + c) + \sqrt{b^2 + (a - c)^2}}{2}},$$

( 3.11 )

$$w = \sqrt{\frac{(a+c) - \sqrt{b^2 + (a-c)^2}}{2}},$$

( 3.12 )

With this information the rotation and scale is determined which can be used to add two more degrees of freedom to the Meanshift method. This results in the Camshift method having three translation tracking and one rotational tracking. The Camshift method was used as one of the tracking method implemented and tested during this work.

## 4.3 Scale-invariant feature transform (SIFT)

The SIFT algorithm was developed by David Lowe in 1999 [27]. This algorithm transforms raw images into feature vectors which will be invariant to translation, rotation and scaling. Locations in the images that are invariant are needed to be identified to produce feature vectors.

According to Lundeberg [28] the only suitable smoothing kernel for scale space analysis is the Gaussian kernel and its derivatives [27]. Rotational invariance is achieved by identifying local maxima and minima of a difference of Gaussian functions applied in scale space. As the 2D Gaussian function can be applied as two 1D Gaussian functions applied in horizontal and vertical directions the 1D Gaussian is

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$$

( 3.13 )

The first input image is convolved with the Gaussian function with sigma of  $\sqrt{2}$  to produce an image A. This is the base level of the pyramid function.

Image A is then convolved with the Gaussian function again with a sigma of  $\sqrt{2}$  to give an effective smoothing of sigma = 2 resulting in image B. The next level of the pyramid function. The difference Gaussian is calculated by subtracting image B from image A [27]. A bilinear interpolation is performed on image B to generate image C with a pixel spacing of 1.5. This is to generate the top pyramid level. The maxima and minima of this function are determined by comparisons between each pixel and its 8 surrounding neighbours at the same pyramid level. When a maxima or minima is detected in a pyramid level, the location of this pixel is determined on a lower pyramid level. If this pixel is still a maxima or minima, the process continues to the bottom, otherwise the pixel is discarded. As most pixels are eliminated with this comparison method, the processing time to determine key features is relatively low. At each key location  $A_{ij}$  on image A, the gradient magnitude  $M_{ij}$  and orientation  $R_{ij}$  are calculated using pixel differences.

$$M_{ij} = \sqrt{(A_{ij} - A_{i+1,j})^2 + (A_{ij} - A_{i,j+1})^2} \quad (3.14)$$

$$A_{ij} = \text{atan2}(A_{ij} - A_{i+1,j}, A_{i,j+1} - A_{ij}) \quad (3.15)$$

The key locations are assigned a rotation by using a histogram of nearby orientation values. This histogram has a Gaussian-weighted window with a sigma of  $\sqrt{18}$ . The peak histogram rotation value is used as rotation at the key location. This method to find the rotation is used as it minimises errors due to lighting or contrast changes [27]. The Sift Key locations, orientations and magnitudes are calculated for any further images. This Sift Key data is used to compare images using a modified k-d tree algorithm known as the best-bin-first search method [29]. Large scale features are also given twice the weighting of smaller scale features, this allows the processing time and accuracy of the algorithm. By determining matches between two images an objects location and orientation can be tracked between images.

#### 4.3.1 Speed Up Robust Features (SURF)

The SURF tracking method is a robust local feature detector based on the SIFT tracking algorithm developed by Herbert Bay et al [30].

The SURF feature detector used the determinant of the Hessian matrix to calculate both the feature location and the scale or magnitude. For a point  $p = (x, y)$  in an image  $I$ , the Hessian matrix  $H(p, \sigma)$  at point  $p$  with scale sigma is defined as

$$h(p, \sigma) = \begin{bmatrix} L_{xx}(p, \sigma) & L_{xy}(p, \sigma) \\ L_{xy}(p, \sigma) & L_{yy}(p, \sigma) \end{bmatrix}, \quad (3.16)$$

Where  $L(p, \sigma)$  is the convolution of the second order derivative of the Gaussian with the image  $I$  at point  $p$ . As a Gaussian filter needs to be discrete and cropped it is simpler to approximate it with box filters. The SURF tracking method was used in this thesis.

## 4.4 Least Squares

Least squares method calculates the difference between two sets of data, a reference data set and a current data set to determine the movement of a 3D object. For this method it is assumed that the patient and/or phantom would only undergo translational motion with no rotational motion. This assumption was made to reduce the complexity of the system.

An initial point cloud is produced and used as a reference for all future point clouds. This point cloud is acquired immediately after the CBCT scan is completed, as the patient is assumed to be in the correct position. This point cloud is known as  $P_0$  is a subset of the whole point cloud acquired in each sensor frame. Searching for the patient in the whole point cloud would be very time consuming so a region which is slightly larger than the point cloud  $P_0$  is used which encompasses  $P_0$ . This extended point cloud is known as  $EP_0$ . In each frame acquired the extended point cloud  $EP_n$  is acquired. The point cloud  $P_n$  is moved in x-y positions to find the best match inside  $EP_n$ . This is illustrated in Figure 4.5.

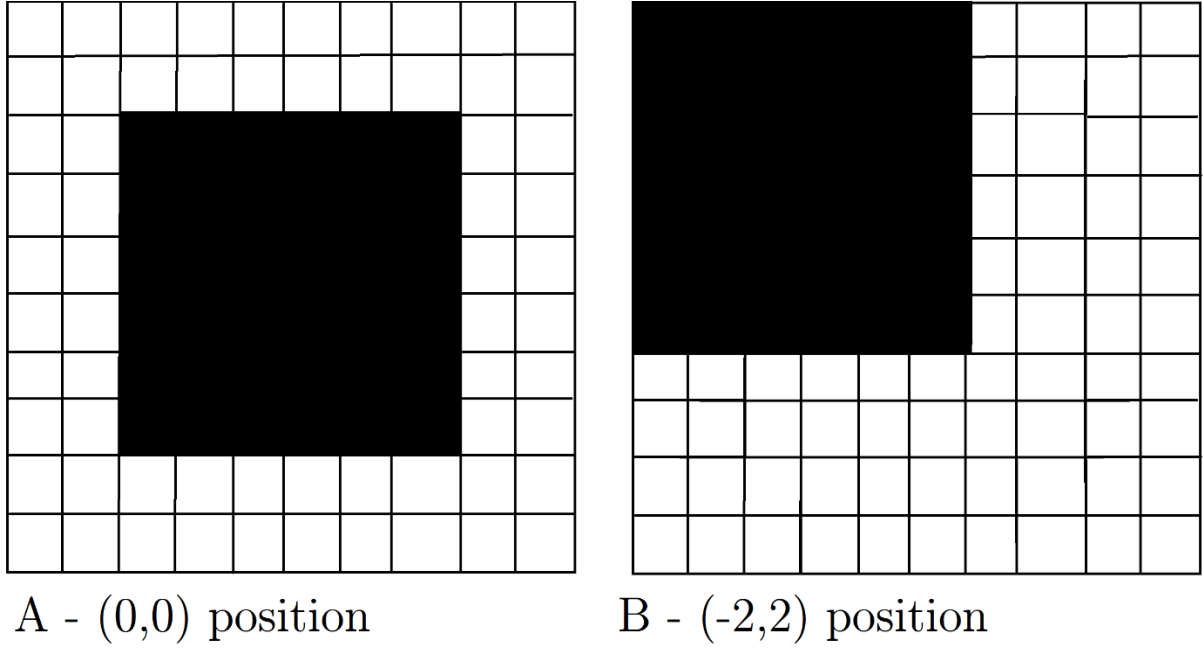


Figure 4.5: Point cloud inside extended point cloud at positions (0,0) and (-2,2)

The point cloud  $P_n$  is moved to every possible position inside  $EP_n$  and at each position the mean difference and standard deviation between  $P_n$  and  $P_0$  is calculated. This is achieved by letting every point in  $P_n$  at position (x,y) consist of an x, y and z position in mm relative to the iso-centre. Then the mean position difference M between  $P_0$  and  $P_n$  is

$$M_x = \frac{\sum_{i=0}^n (x_{ni} - x_{0i})}{n}, M_y = \frac{\sum_{i=0}^n (y_{1n} - y_{0i})}{n}, M_z = \frac{\sum_{i=0}^n (z_{1n} - z_{0i})}{n}. \quad (3.13)$$

And the Standard deviation for this position is given as

$$\sigma = \sqrt{E([X - M]^2)}, \quad (3.14)$$

Where the E operator is the average of  $[X - M]^2$ , where X contains all different points in the two point clouds such that  $X_i = P_{1i} - P_{0i}$ . The (x,y) position of  $P_n$  with the lowest  $\sigma$  value is the best match to the initial point cloud  $P_0$ , thus this must be where the patient has moved, and the resulting mean difference M is the movement of the patient relative to the initial position. All positions are not searched as it is assumed that the phantom or patient should not be moving very fast.

Due to the high frame rate of the sensor at a distance of 3 metres an object would need to be traveling faster than 0.64m/s for the tracking system to fail when tracking the object. This method is one of the tracking methods implemented and tested when developing the system.

## 4.5 Iterative closest point (ICP)

The ICP method is an extension on the least squares method that allows mapping of full 3D objects in six degrees of freedom (6DOF). The first 6DOF algorithm was proposed by Paul Besl and Neil McKay in "A Method for Registration of 3-D Shapes" [31]. This algorithm is designed to align two sets of data, in this projects case, two point clouds.

Let a point in 3 Dimensions be  $\vec{r}_k = (x_k, y_k, z_k)$  then the distance between two points is

$$\|\vec{r}_2 - \vec{r}_1\| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (4.15)$$

Let A be a point cloud with  $N_A$  points such that  $A = \{\vec{a}_i\}$  for  $i = 1, \dots, N_A$  therefore the distance between the point  $\vec{r}_p$  and the point cloud A is



$$d(\vec{r}_p, A) = \min_{i \in \{1, \dots, N_a\}} d(\vec{r}_p, \vec{a}_i) \quad (4.16)$$

Let there be a line  $l$  which connects points  $\vec{r}_1$  and  $\vec{r}_2$ . The distance between this line and  $\vec{r}_p$  is therefore

$$d(\vec{r}_p, l) = \min_{u+v=1} \|u\vec{r}_1 + v\vec{r}_2 - \vec{r}_p\| \quad (4.17)$$

This is done by comparing each point to every other point. This function is similar to the least square function, due to going from an  $n$  function to an  $n$  squared function, where  $n$  is the number of points, the computational time to process a single iteration is very long. The advantage of this method is it is able to perform a six degree ridged transformation between two point clouds or a point cloud and a plane.

Due to the slow speed of this method of approximately 1 second to process a position, this method was not used for tracking. However, it was used for determining the translation between the sensor co-ordinate system and the LINAC Isocentre.

## 4.6 Kinect Fusion

Kinect fusion is a project developed by Shahram Izadi et al [32] at Microsoft™ Research, Cambridge, United Kingdom. This project demonstrates a real-time 3D reconstruction and interaction using a Microsoft™ Kinect v1 sensor. While the primary purpose of the Kinect fusion project is for mapping and scanning fixed 3D environments with a moving sensor, it is possible to be modified to enable a fixed sensor to track a moving object. Kinect Fusion solely uses depth data obtained from the Kinect v1 for tracking and 3D surface mapping. The sensor tracking implementation which is used to track the relative position of the sensor is based on the ICP method discussed in the previous section. In Kinect fusion a GPU execution of ICP has been included. This is to ensure this algorithm runs at a high speed. The Kinect fusion algorithm is illustrated in Figure 4.6 where the raw data has been obtained from a depth sensor in this case a Kinect v1 sensor. The raw depth is converted to a depth map which is used as input for the Sensor tracking algorithm which finds the 6 degree of freedom (6DOF) pose and removes any outliers.

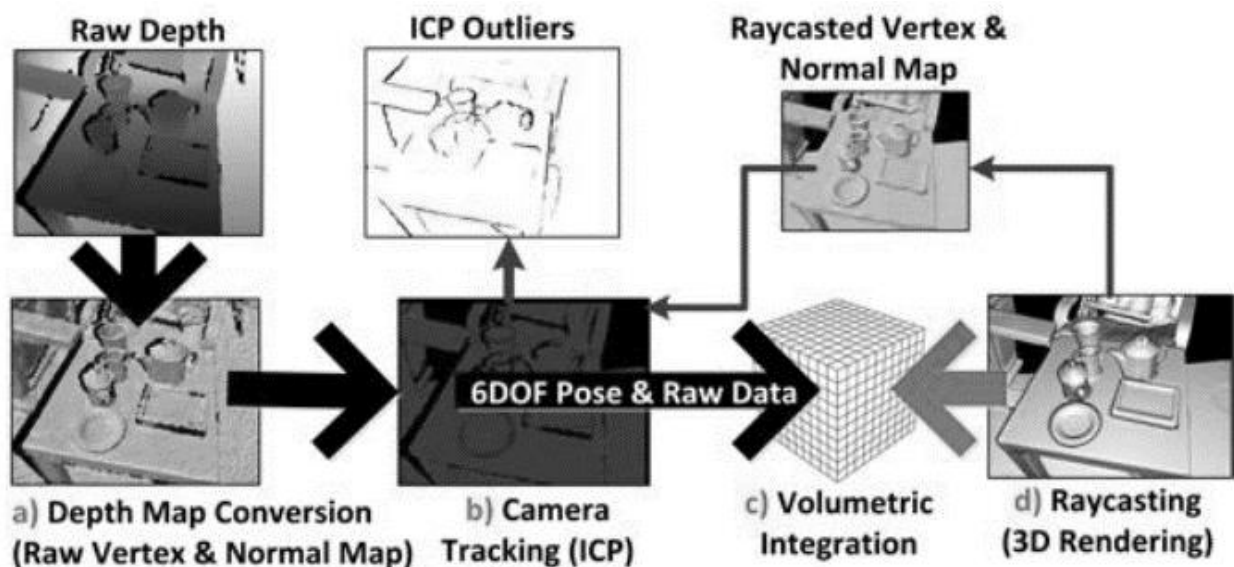


Figure 4.6: Kinect Fusion reconstruction and pipeline illustrating the order of operation [19]

A 6DOF pose is the three current rotation and translations which the sensor or object has been moved by from the initial position. The raw depth data with outliers is integrated into a voxel based 3D volume which is updated with each new frame. This volume will slowly build a 3D reconstruction of the scanned object. The two point clouds which ICP compares are the initial input point cloud from the raw data and a ray casted point cloud of the current 3D model from the point of view of the last known position. The Kinect fusion algorithm was modified for use in this project, the modifications change the system, and so only an initial volumetric integration is preformed to generate the 3D voxel based volume. The sensor's relative position to the initial position is tracked and from this information the real movement of the 3D surface can be calculated.

# Chapter 5: Software

---

In this project an application was developed to collect data from the Kinect sensors and to process this data in real time, with the purpose of determining and detecting patient motion from an initial position.

This chapter introduces the software tools utilised throughout this project and the decisions behind the use of these tools. This chapter will also discuss the design of the Graphical User Interface (GUI), the coordinate systems used, and the process of mapping between coordinate systems.

## 5.1 Development Environment

As a software solution for the needs of this project does not currently exist an extensive component of the work done was in the development of a dedicated PC based application. During the initial design phase, decisions had to be made concerning the appropriate language and development environment. As discussed in Chapter 3 Kinect v1 and v2 sensors were chosen for the depth sensor, and because of this decision there exist certain software constraints which have been introduced by Microsoft™. The software constraints were primarily based on the software development kit (SDK) for the Kinect v2 device. These requirements necessitated that Windows 8 or higher be used and Microsoft™ Visual Studio 2012 must be used as the interactive development environment (IDE). The remaining choice of programming language was also confined to either C# or C++. From these options it was decided that C# would be used for the entire coding section of the project, as C# is simple to use and can produce working graphical user interfaces easily.

### 5.1.1 Programing libraries and Software Development Kit (SDK)

The image processing toolkit, Open Source Compter Vision (OpenCV) was used throughout this project. OpenCV was initially developed by the Intel Corporation and is now currently supported by Willow Garage and Itseeez. A software toolkit is a collection of prewritten functions which can be implemented into any program.

The purpose of a toolkit is to increase coding productivity by providing prebuild functions that are commonly used rather than having everyone code them from scratch. This toolkit provides the Mean shift, Camshift and Surf tracking functions used throughout this project, in addition to a large range of image processing functions. OpenCV is a C/C++ Library and therefore required a wrapper to enable the use of OpenCV functions in C#. For this purpose EMGU CV was used, this allows the OpenCV functions which are C/C++ functions to run and be called from a C# environment.

Interfacing with the Kinect sensors was achieved through the Kinect for Windows SDK. Both Kinects outputs frames at a rate of 30Hz. Each frames consist of a depth image and a Red, Green, Blue, and Alpha (RGBA) image. The (RBGA) image is in the format of a Byte array of size 1,228,800 for the Kinect v1 and 8,294,400 for the Kinect v2. The depth data is in the format of a Short array of length 307,200 bytes for the Kinect v1 and 217,088 bytes for the Kinect v2. After a frame is received further processing can be performed this processing should complete before the next frame arrives. The SDK for the Kinect v1 also provides access to the Kinect fusion functions used for tracking. Kinect fusion as discussed in chapter 3 has yet to be implemented for the Kinect v2 SDK.

## 5.2 Graphical User Interface (GUI)

One of the design considerations was to make the user interface as user friendly as possible. One of the motivating factors for this is that ultimately the software would be used by radiation therapists in the clinical environment. Therapists when delivering a radiotherapy treatment already have a large number of screens to monitor. To ensure minimal invasion of the therapist's working environment, the GUI was designed with user simplicity in mind. The GUI design had the following aims; Easy to read important information available at a glance, minimalistic design so inexperienced users do not have difficulty interacting with the system, colour coded information to inform therapists when a patient has moved out of position.



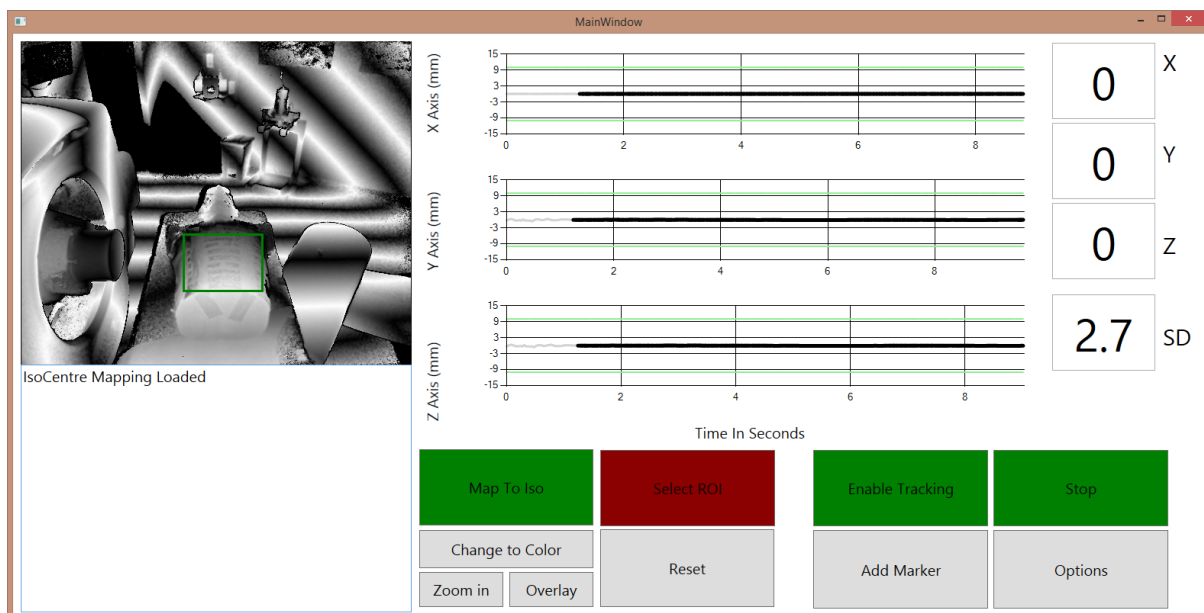


Figure 5.2: Screenshot of the current GUI Design tracking a stationary phantom at the Proton therapy centre in Seattle

Figure 5.2 shows the four primary sections were decided on to be incorporated into the GUI; a live stream from the sensor as seen in the top left, large x, y, z average position readouts seen in the right hand side of the GUI, x, y, z graphs of current position as a function of time in the central section, and a control section seen at the bottom of the GUI. The graphs and position readout were designed to be quite large so they could be easily read when glancing at the GUI. This is the important information generated by the system for the therapist. Figure 5.2 displays the GUI monitoring a torso phantom in the

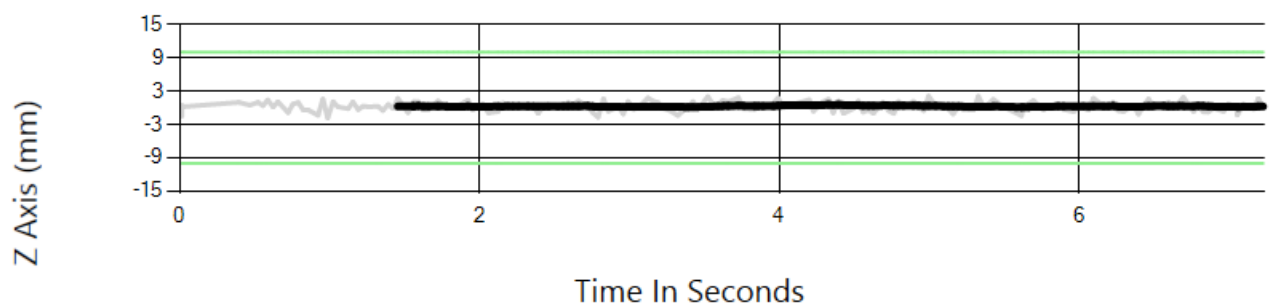


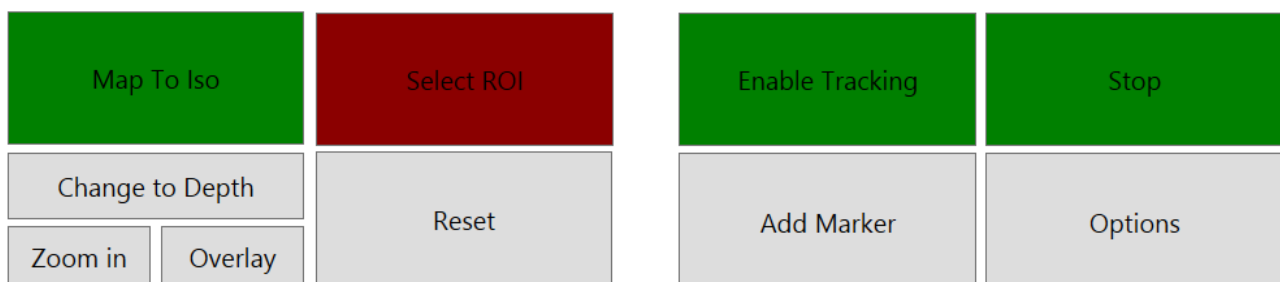
Figure 5.1: Graph component of the GUI

Proton therapy centre. The viewer is shown running in infra-red mode, with a grey scale

depth image. The graphs are showing movements from the initial position in mm as a function of time, currently the GUI is displaying no movement of the phantom which is to be expected.

Figure 5.1 displays the components of the graphs. Each graph has three components: a thin grey line which is the current position at the current time, a bold black line which displays the moving average position at the current time and two green lines which are the set tolerance values for this setup.

As can be seen from the GUI for that particular situation they were set to 10mm. The moving average determines the position by averaging the last 30 values. The position readouts are displayed on the top right. These readouts show the moving average position



*Figure 5.3: Control section of GUI*

in mm as discuss earlier, and the standard deviation of the current position compared to the original surface.

Figure 5.1 displays the controls section located in the bottom right section of the GUI. These provide the following control options;

- Map to Iso; will perform a basic rotation mapping between the LINAC Isocentre and the sensor's isocentre, this process will be discussed latter in the chapter. This alignment will usually be performed when the sensor is installed and during a monthly check. This alignment is saved for all treatments.
- Select ROI; allows the user to select a Region of Interest (ROI) which they intend to track during treatment. This is achieved by clicking on two corner points of a rectangle on the viewer screen with the mouse pointer.
- Enable Tracking; allows the user to enable or disable the tracking software, which under normal operation will remain on.
- STOP/START; starts and stops the software.
- Change to colour; (Change to depth), can change the main viewer to an RGB colour view or Greyscale depth view as shown in Figure 5.2.
- Zoom in; scales the ROI to fit to the entire viewer window.
- Overlay; enables a colour overlay which shows colour coding based on the distance between the surfaces. No colour if they are an equivalent distance, Red if the distance is negative, and blue if the distance is positive. This is illustrated in Figure 5.4 where a pseudo patient sat up from a lying down position, resulting in the ROI changing to a red colour.
- Reset; removes all graph data and the initial surface, resetting the program to the current position.
- Add marker; allows the user to place a marker in the graph series for later reference. For example, this can be used to identify when the beam is turned on.
- Options; brings up an options GUI which is currently used for system development and testing.

The GUI also has a warning system which changes the colour of various regions to red if the patient moves significantly from the initial position. This is demonstrated in Figure 5.4, where the patient has moved out of position in the z direction by  $-22\text{mm}$ , such a large shift is significantly more than the  $10\text{mm}$  tolerance. The box and the numerical display have therefore changed colour to better alert staff monitoring treatment. The area in the bottom left is used for system messages shown in Figure 5.2 and Figure 5.4.

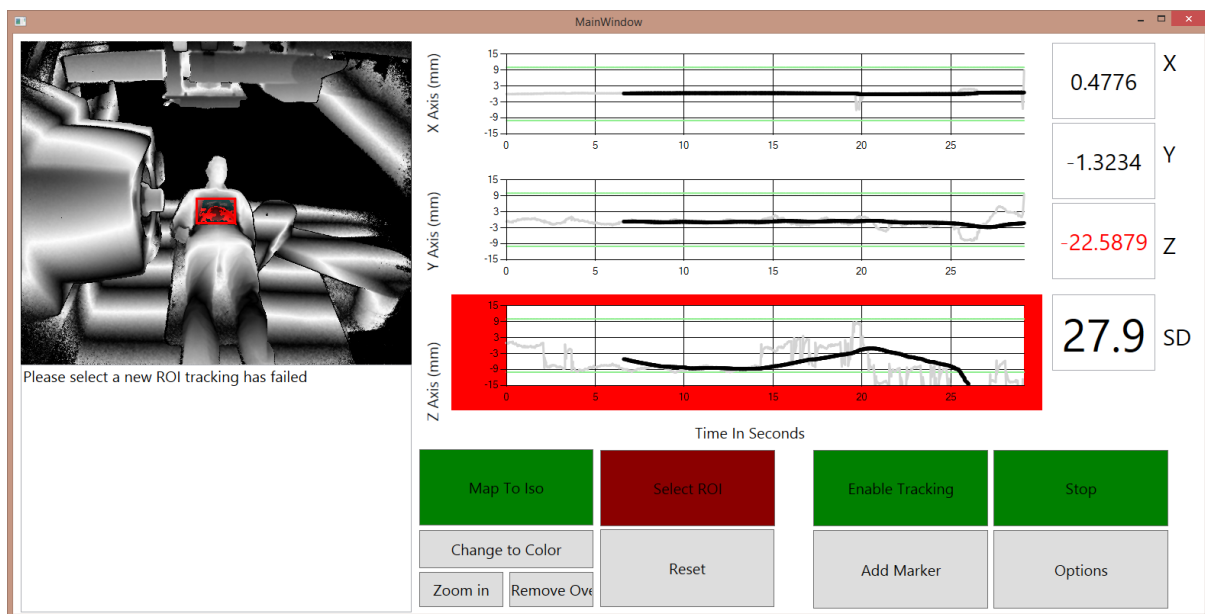


Figure 5.4: Screenshot of the current GUI Design tracking a volunteer who has sat up during tracking demonstrating the colour warning system

## 5.3 Parallelisation

Each frame has 300,000 or 200,000 pixels from the Kinect v1 or Kinect v2 respectively. Therefore a significant amount of processing power is required for tracking, coordinate system mapping, and acquiring position measurements.

As the frame rate of the Sensor is 30Hz, each frame must be processed in less than 0.033 seconds to ensure the system is in a position to be able to process the next frame when it becomes available. Parallelisation of intensive sections of code was required to accomplish this. Parallelisation is where instead of processing once piece of work at a time multiple processes can run at the same time using multiple CPU cores. To achieve this the System.Windows.Threading library was utilised to improve performance for computationally intensive sections of code. This library is a standard C#.NET library developed by Microsoft™. The primary function which was used from the Threading library was the Parallel.For loop, which can parallelise any For loop as each loop cycle is independent from the total loop cycle. This function automatically handles thread syncing and is capable of utilising a large number of real and hyper-threaded threads.

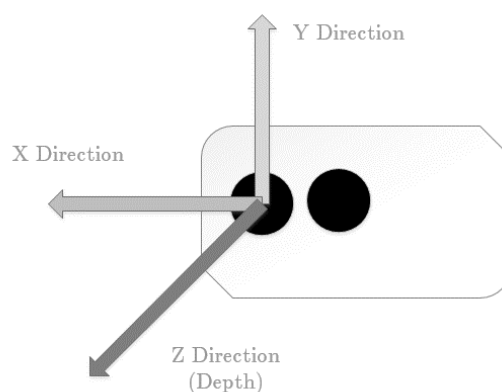
## 5.4 Coordinate systems

The use of coordinate systems in Radiotherapy is essential to ensure that a patient's position is well known. In addition to these coordinate systems an additional coordinate system is introduced from the Kinect sensors. This section will discuss the various sensor coordinate systems used in this work.

### 5.4.1 Kinect Sensors co-ordinates

The Kinects operates a left handed co-ordinate system originating at the Kinect sensor. Data from the Kinect v1 and v2 is reported in millimetres (mm) is shown in Figure 5.5 in which the axis have been labelled X, Y and Z. Both the Kinect v1 and v2 utilise the same coordinate system. As discussed in chapter 3 and 4 the Kinects only receives information about the depth of a pixel and the (x, y) location must be inferred from the pixel's position.

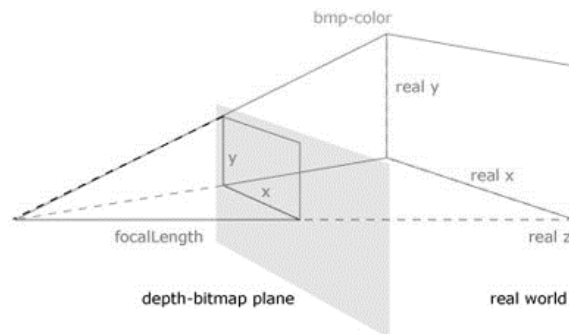
This is achieved by using the known field of view (FOV) of the sensor and the depth of the pixel. Using this information, a simple trigonometric calculation is preformed to find the relative position of that pixel. This calculation is performed by an in-built function in the Kinect for Windows SDK, the resulting data is stored in an X, Y, and Z point cloud.



*Figure 5.5: Illustration of the Kinect coordinate system, origin is located at the depth camera in the Kinect sensor.*

### 5.4.2 LINAC couch coordinate system

The LINAC couch system used in radiotherapy has 3 rotation axes and 3 translation axes. This is illustrated in Figure 5.7. LINACs are capable of 2 forms of movement, gantry rotation and head rotation, around a fixed point in space. This point is known as the



*Figure 5.6: Illustration of trigonometric calculation to determine the real position of pixel using the position of pixels in the bitmap [44]*

mechanical isocentre, and this isocentre is ensured to coincide with the radiation isocentre. The couch is capable of 4 movement forms, couch rotation around the mechanical isocentre and couch translation in 3 directions. The 3 movement vectors are shown in Figure 5.7 as X, Y, and Z and are known as latitude, longitude and vertical respectively. The origin for the couch translation coordinates are located at the geometric isocentre. The Head, Couch and Gantry are shown positioned at 0 degrees in Figure 5.7. As the coordinates used by the Kinect sensors do not coincide with the coordinate system used in radiation oncology, a transformation is applied so that information obtained from the patient monitoring system is useful. The next section will discuss this transformation.

## 5.5 Kinect to LINAC coordinate transformation.

As discussed in the previous section/chapter the Kinect sensors returns data defined relative to its own coordinate system and a transformation is required to map acquired data into the LINAC table coordinate system. Mapping between these coordinate systems is achieved with a rigid transformation as there is a one to one mapping between the coordinate systems.

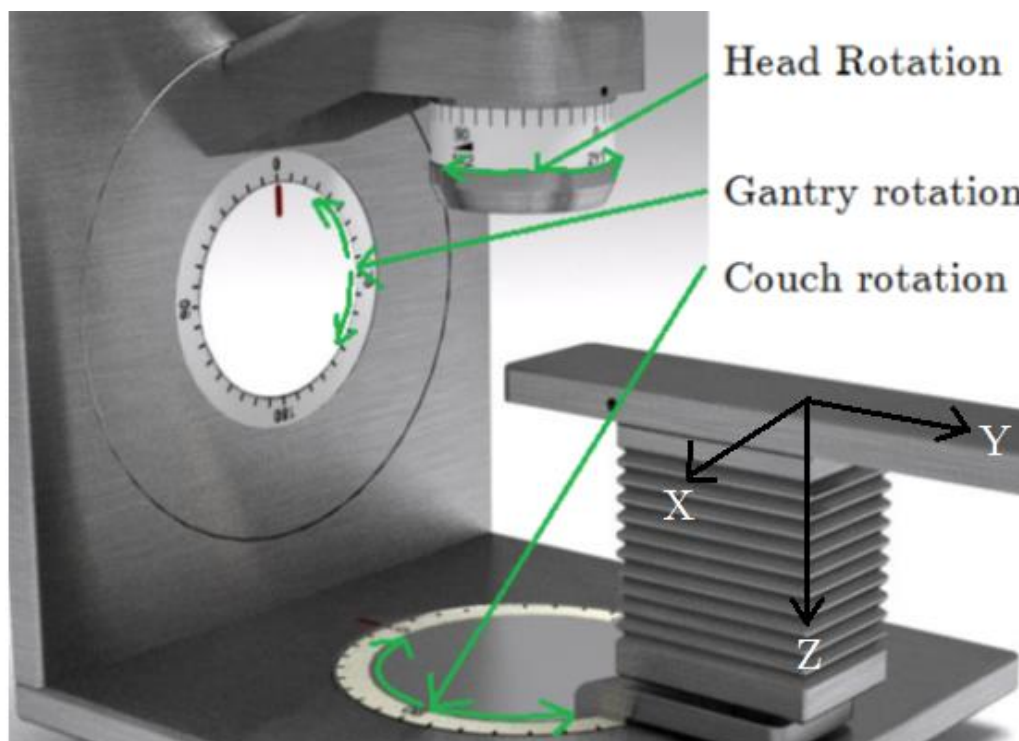
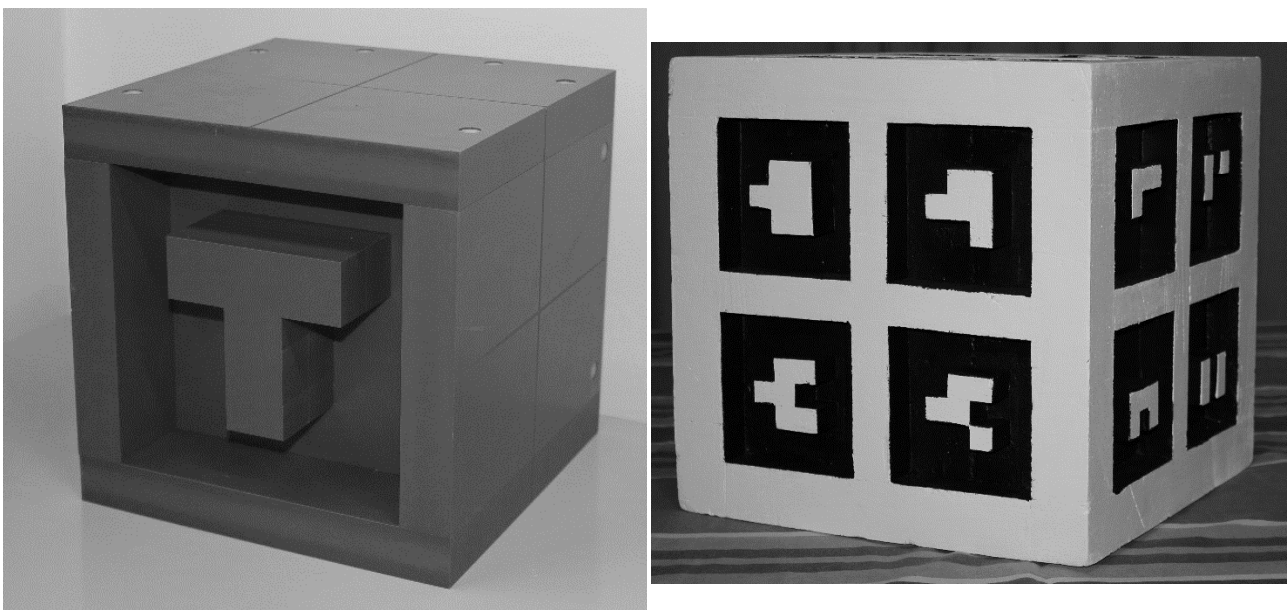


Figure 5.7: Model of LINAC and couch movements and coordinates [38]



Since both coordinate systems use metric units, only a translation and a rotation was required. To facilitate this transformation, the phantoms shown in Figure 5.8 was made.

This phantom, referred to as the Isocentre Cube (left image), was designed specifically for this project. An older design was used while at the University of Washington (right image). The shape was designed to be a cube, to ease alignment when positioning at the LINAC Isocentre. Grooves are cut along the sides of the Isocentre Cube so that the room Lasers can be used to achieve an accurate alignment in a LINAC bunker room. These grooves can be seen on the top and side of the cube in Figure 5.8. The phantom

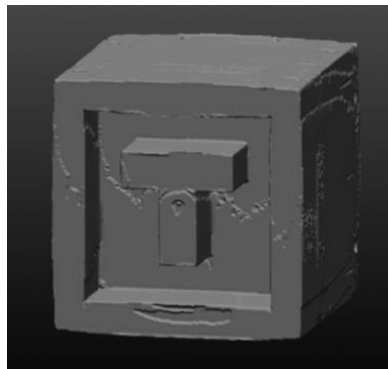


*Figure 5.8: Image of the Isocentre Cubes. University of Canterbury cube (left), University of Washington cube (right).*

also required a non-flat rotationally invariant front surface so that the Kinect sensors are able to uniquely identify the surface. This is the “T” on the front of the cube seen in Figure 5.8.

The cube was built out of a PVC plastic to ensure it would not warp or deform. This material must also be able to be scanned with a CT scanner. The coordinate transformation requires a reference 3D model of the Isocentre cube. This was obtained from a CT scan from the Christchurch Hospital. The cube's surface was then contoured and an output of these contours was converted to an appropriate 3D format. Figure 5.9 shows the 3D reference model obtained from the CT scanner.

The model has some minor surface artefacts which are from the CT scan. However, these artefacts will have no impact on any transformation as they are only superficial with only a very small deviation from the correct surface.



*Figure 5.9: Image of the 3D model of the Isocentre cube produced from the CT scan*

With this information a mapping can be performed between the reference 3D model and a point cloud as discussed earlier in this chapter obtained of the cube from the Kinects. As the cube is positioned at the LINAC isocentre, the point cloud from this reference scan is also located at the Isocentre of the LINAC coordinate system. The point cloud obtained from the Kinect sensors is cropped to ensure that only the tracking cube phantom is visible. A two-step method was developed for this project with the first step being an initial alignment. This is achieved by having the user select 4 corner points from the front face of the phantom. To remove noise and possible errors, 25 neighbouring points encompassing the selected corner, are averaged to produce a more reliable position. These points are then aligned to four corner points of the Isocentre cube, which were obtained from the CT scan. This process is achieved by finding the average centre of the points and translating them so the difference between the centres is zero. The rotation transformation is then calculated. This is done in 3 steps, by processing one rotation axis at a time. The combination of these rotations result in the rotation matrix and the resulting translation vector is used to produce a 4 by 4 transformation matrix. A 4 by 4 transformation matrix consists of the rotation matrix in the top left 3x3 section and the translation vector in the right side of the matrix. The Kinects uses the new transformation matrix and transforms the obtained point cloud so it aligns with the isocentre. This first step gives a good initial alignment and fundamentally the translation is correct. The rotation transformation is not yet accurate enough for this project as it can be out by up to 8 degrees, therefore a second step is required. The second step is a fine alignment using the iterative closest point method which matches the initial transformed point cloud with the reference 3D model of the cube. Once this step is complete a transformation matrix has been defined and all

data from the sensor can be accurately mapped into LINAC coordinates with less than 0.5 degrees of error in the rotation.

This chapter has discussed the software libraries to develop a system with a GUI running in LINAC coordinates. Chapter 6 will discuss testing of the various components of the system introduced in the last few chapters.

# Chapter 6: Characterisation

---

This chapter discusses the various tests applied to characterise the system's performance during this thesis. As the system is reasonably complex and consists of many components, testing of each component needs to be completed individually. However some components are dependent on other components, therefore the independent components are tested first to ensure components built on top are not effect by the underlying functions. This requires that results from the first tests were used for subsequent tests. This chapter is divided into three sections. The first will discuss the methodology of how the system was developed. The second section shows results from testing the hardware components, and the final section discusses the software components of the system.

## 6.1 Sensor placement and setup

### methodology

The following Equipment was used throughout the project:

- Kinect for Windows v1 Camera, with a viewable range of 800–4000mm
- Kinect for Windows v2 Camera (USA model) with 230Volt to 110Volt transformer and a viewable range of 500–4000mm.
- Dell Inspiron 7520 Laptop with an Intel Core i7 3632QM 2.20Ghz quad core Central Processing Unit (CPU), 8GB of Random Access Memory (RAM), AMD 7730M Graphics Processing Unit GPU and Universal Serial Bus (USB) 3.0 Support.

- Adjustable Tripod
- Tracking Cube v1 and v2
- Dynamic motion platform (UW)
- Dynamic motion platform (UC)

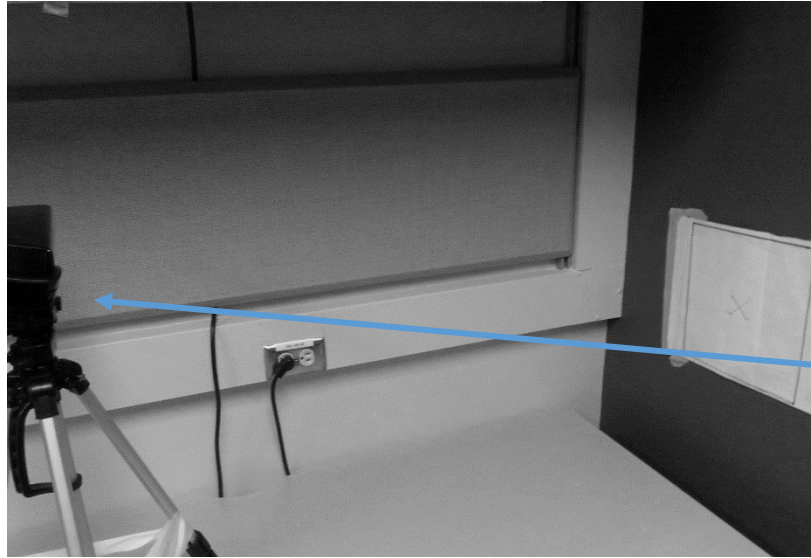
To ensure test repeatability, the following two conditions were maintained throughout all tests: Kinect cameras were kept parallel to the ground and an operating range of 850–3000m was maintained for both Kinect cameras. Keeping the Kinects parallel to the ground is to reduce any errors produced from co-ordinate system transforms. This operating range was chosen to ensure both cameras functioned correctly, and neither was operating at their minimum or maximum range.

## 6.2 Hardware

This section analyses the Kinect v1 and the Kinect v2 hardware characteristics by testing both cameras with the same software at the same time. Non-disclosure act (NDA) disclosure: Throughout this project we have been part of the Kinect for window v2 alpha program and therefore have been given an alpha Kinect v2 sensor and software from Microsoft™. While this sensor is likely to be very similar to the final released sensor, it is still in the development phase and all results may change. Hardware testing of the Kinects consists of an electronic stability test and a depth accuracy test, the section then continue to discuss hardware limitations of the Kinect sensors.

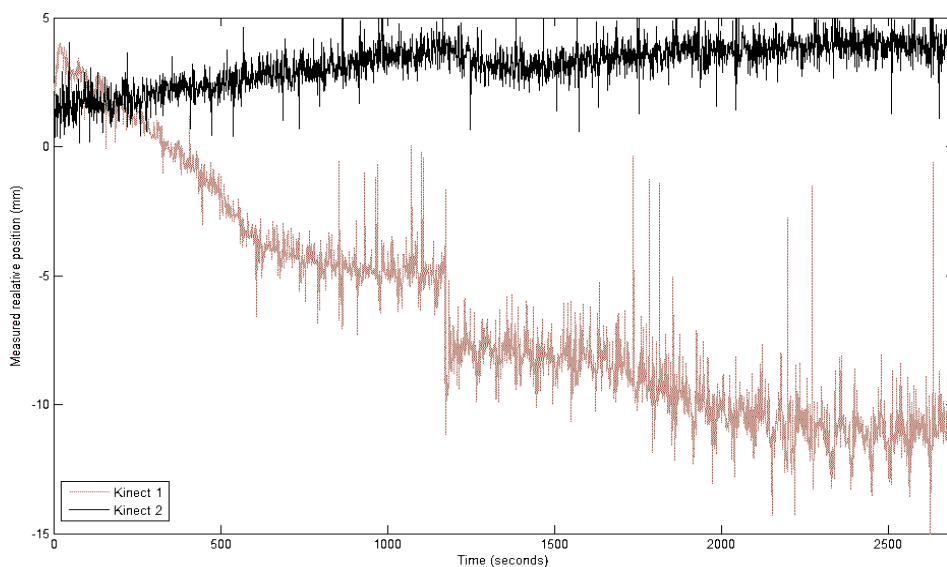
### 6.2.1 Kinect Long term stability

The electronic stability of the Kinect sensors over an entire clinical day is highly important if these sensors are to be used in a clinical environment. It has been suggested that the most optimal use of radiotherapy machines is a two shift program operating from 6am till 8pm [4]. Therefore, the camera system must be stable and reliable for the entire 14 hour time period. A test was devised to establish the long term stability of both cameras. The setup is shown in Figure 6.2. Both Cameras were placed in a closed office room with no internal airflow. The office room was also shut off from external light sources and only used internal lighting. This was to ensure Infrared radiation from sunlight did not interfere with the experiment. The two Kinect cameras were mounted on a tripod, and positioned 1.5m from a wall as shown in Figure 6.2. Markers were placed on the wall to ensure both cameras were observing the same approximately  $1\text{m}^2$  section. All tracking software was disabled for this experiment.



*Figure 6.2: Image showing the Kinect stability testing setup, displaying the position the sensor was placed and the size of the area used for testing*

Figure 6.3 shows data obtained over a 17.5 hour period from both the Kinect v1 and Kinect v2 cameras. Both Kinect cameras were monitoring a fixed area of smooth and uniform wall. Data shown in Figure 6.3 established that there was an initial warm up period for both cameras of around 40 minutes. However, once this initial period had

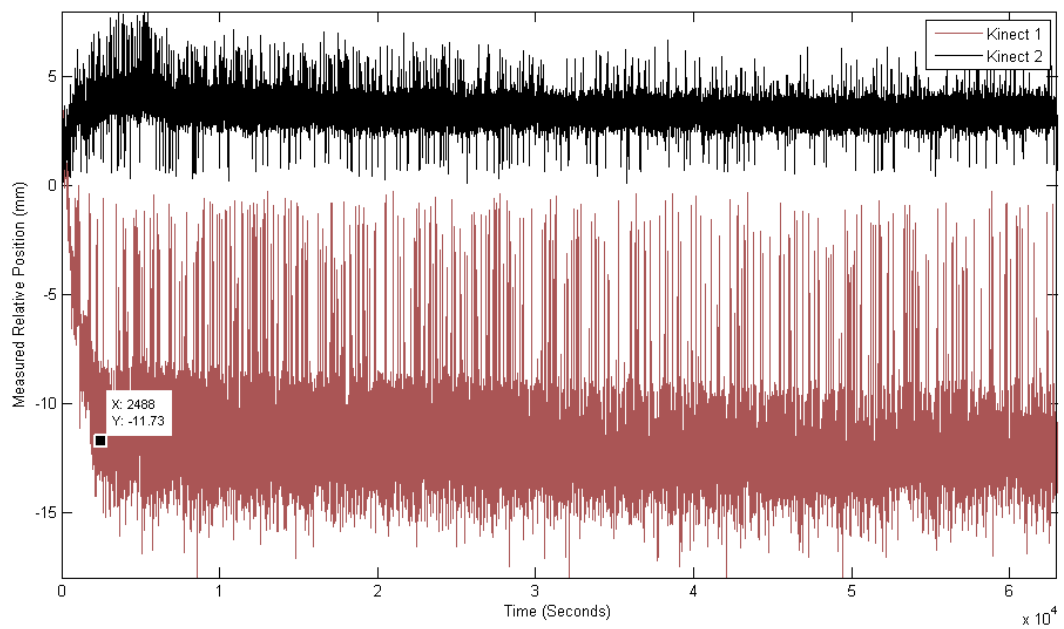


*Figure 6.1: Electronic stability of the Kinect v1 (red) and v2 (black) demonstrating the warmup period*

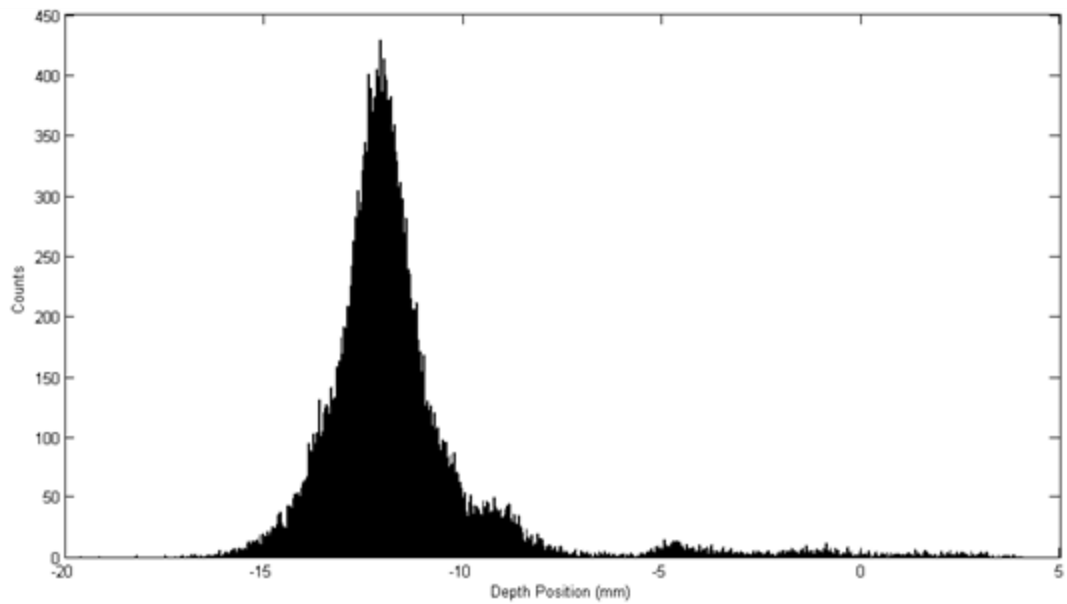


passed, the recorded data were very stable over the remainder of the test. A magnified section of this warm up region is shown in Figure 6.1.

This initial position of zero is not able to be shown as there is a 30 second start up time while the software obtains an initial reference surface. The timer does not start until after this 30 second period resulting in the graph starting at zero. From the initial position the Kinect v1 moves rapidly to a value of 5mm, 20 seconds after initial start-up. A discontinuity is observed in Figure 6.1 at time 1100 seconds the Kinect v1 demonstrates a sudden downwards drop, the cause of this is currently unknown however this does not impact the overall trend. A downwards trend is then observed over the next 40 minutes or 2400 seconds, until a stable position of -11.5mm from the initial position is reached. Similarly the Kinect v2 shifts rapidly to a value of 0.75mm 30 seconds after initial start-up. A slight upward trend is then observed over the next 40 minutes or 2400 seconds, until a stable position of +3.5mm from the initial position is reached. The results of the data shown in Figure 6.3 demonstrate that both cameras are able to remain stable throughout an entire clinical workday. The only complication being, the camera will need to be turned on at least 40 minutes before the first treatment or the sensor could be left on permanently however the impact of leaving the sensor on for more than 24 hours has not been studied as of yet. This could possibly be included in the morning QA that hospitals perform. For all future tests, the Kinect cameras were allowed a warm up period of 40 minutes.

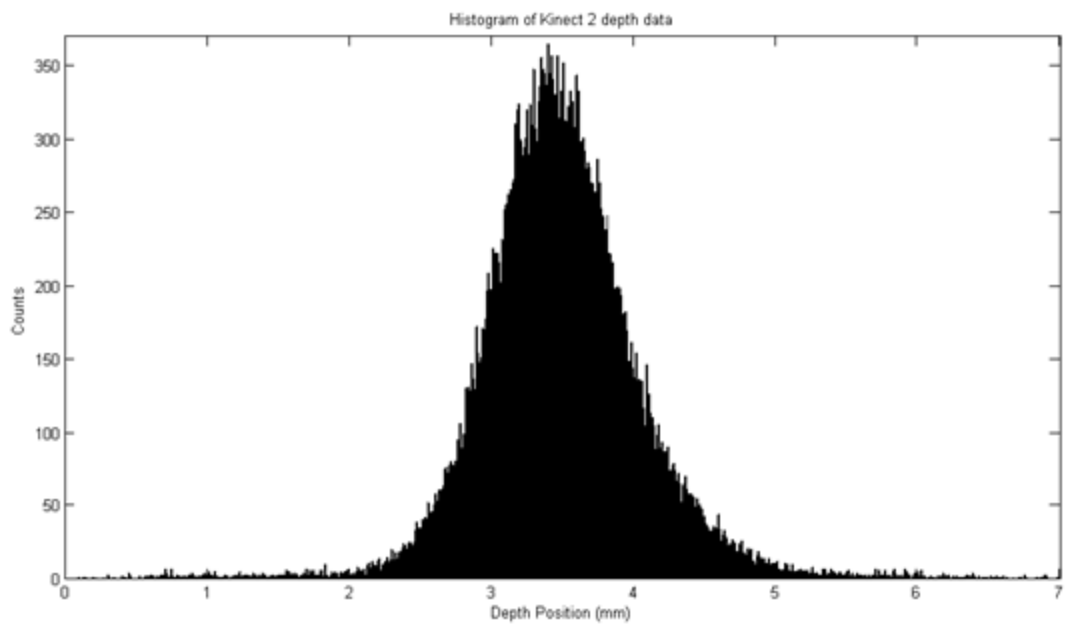


*Figure 6.3: Electronic stability of the Kinect v1 (red) and Kinect v2 (black) over a long period of time demonstrating the warm up period observed.*



*Figure 6.5: Histogram of Electronic stability over period of time, from the Kinect v1 depth data*

Figure 6.5 and Figure 6.4 are histograms of the long term data, note that Figure 6.4 has a significantly smaller scale. The data displayed is the relative averaged measured



*Figure 6.4: Histogram of Electronic stability over period of time, from the Kinect v2 depth data*

position of all depth points. Since most of this data is true random noise, it follows a Gaussian distribution.

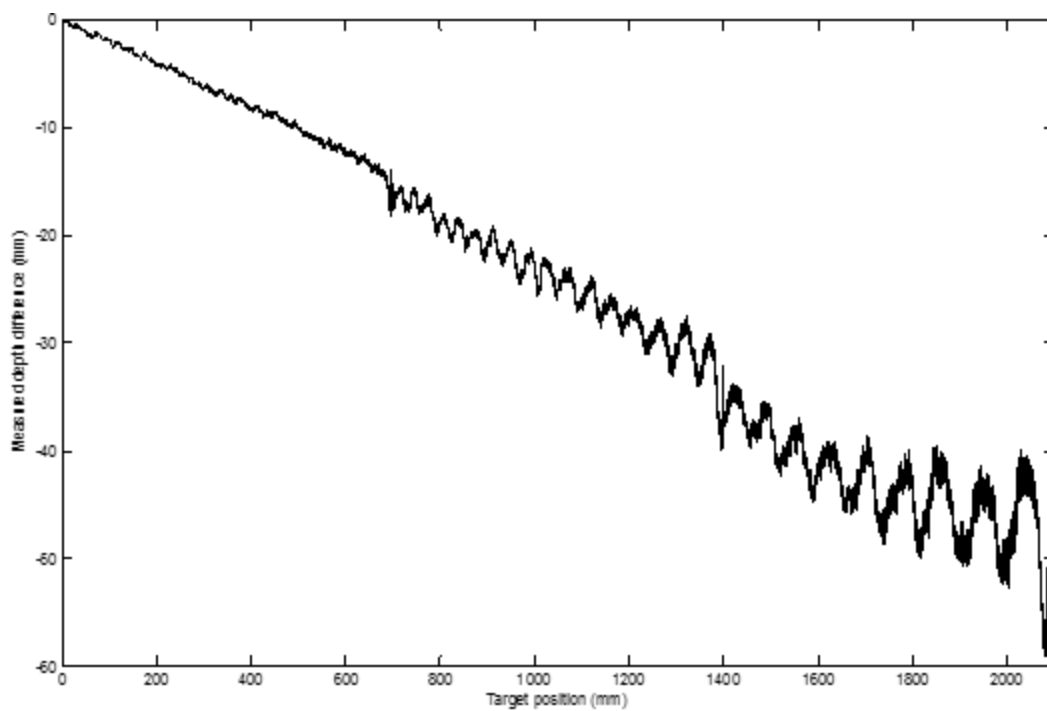
The Kinect v1 has a mean position of  $-11.5\text{mm}$  and the Kinect v2 has a mean position of  $3.5\text{mm}$ . As the figures demonstrate, the mean position is constant over time. However, there is a large amount of noise in the system. The standard deviation of the Kinect v1 is  $2.4\text{mm}$ , while the standard deviation of the Kinect v2 is only  $0.6\text{mm}$ . This clearly demonstrates that the Kinect v2 has four times less variation. The mean position is not zero for this system, as it is a position relative to the initial position. David Fiedler et al [33] in their paper "Impact of Thermal and Environmental Conditions on the Kinect Sensor" demonstrated that the Kinect v1 sensor has a warmup period and recommended that the sensor is left to warm up for 60 minutes to remove any temperature induced warmup errors. This was confirmed with our results and therefore to ensure that each sensor was fully warmed up before each test they were turned on 60 minutes prior to use.

### 6.2.2 Kinect Depth resolution

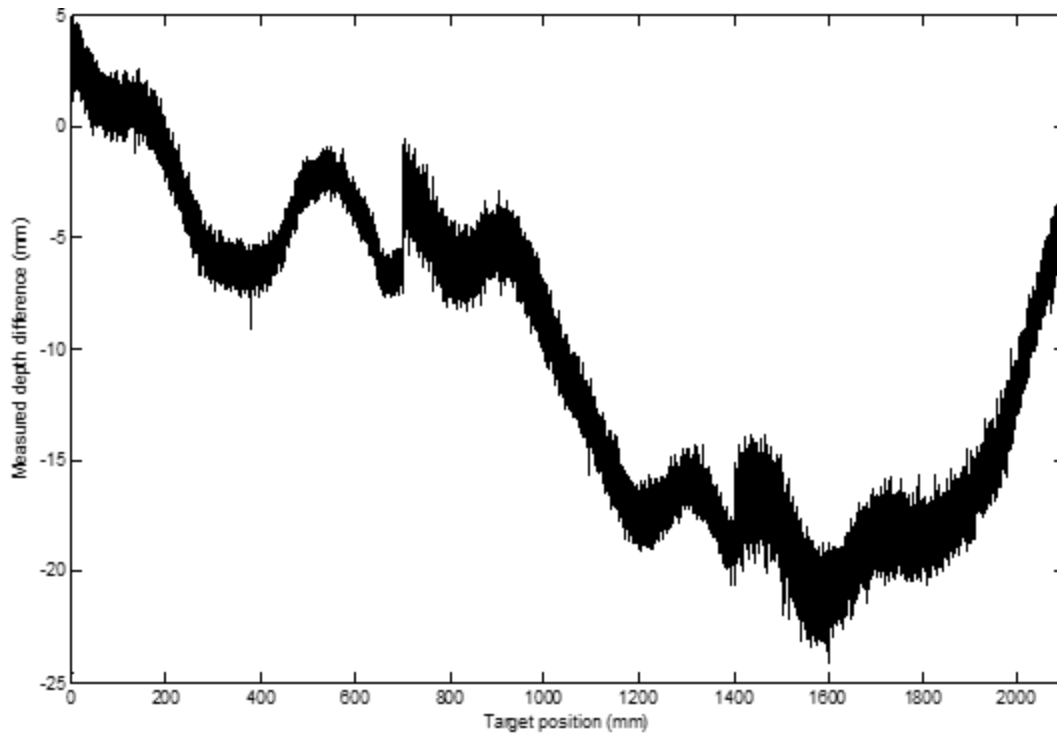
A test was devised to determine the accuracy of the depth measurements that the Kinect cameras output. Using the known accuracy of a milling machine, a target object could be moved toward the Kinect at a predefined rate. A smooth flat target object was attached to the milling machine table. The target was then moved toward the Kinect camera at a rate of 10mm per minute over a range of 700mm. Both Kinect cameras were placed at a distance of 850mm from the edge of the milling machine, as this is the minimum distance a depth measurement can be made. The Milling machine only had a range of 700mm, therefore the Kinect cameras were moved backwards from an initial position of 850mm to  $850+700\text{mm}$  then  $850+1400\text{mm}$ . It was decided that the target should move towards the cameras rather than going away, as this would reduce the chance of the ROI selected moving off the target. All tracking was disabled for the test as this test was to determine the accuracy of the raw depth data. The Iso-transform matrix was set to the Identity matrix to ensure all movement was only the depth direction of the sensor. The cameras were aligned so they were parallel with the milling machine and in direction of travel. This was achieved with surveying lasers and a spirit level. Both cameras were given a warm up period of 90 minutes to ensure no inaccuracies from warmup affected the test.

Figure 34 displays the difference graph between predicted position and position obtained from the Kinect v1. This graph should be a flat line at 0mm over the entire range of target positions from 0mm to 2100mm. However, the data displays that the depth scaling calculation is off by 25mm per meter. This is a 2.5% error in the depth measurement, and should have a minimal effect on measurements of short ranges however at larger distances this is likely to be an issue. If the effect is repeatable then software calibration may be implemented to correct this error. However, further data analysis would be required for this to be performed on the Kinect v1 camera. The data also shows a sinusoidal function with increasing intensity as distance increases. This is presumed to be an artefact of the depth sensing method which the Kinect camera uses. The sinusoid has minimal effect at small depths. However, the impact increases as the depth increases. Therefore the Kinect v1 should be operated as around a distance of 1m from the patient, as this is the minimum reasonable distance.

Figure 6.7 displays the data obtained from the Kinect v2. This data does not have an overall trend but has a semi random pattern. It is unknown why the Kinect v2 produced this pattern.



*Figure 6.6: Difference graph between Kinect v1 depth data and the correct position demonstrating the accuracy over distance*



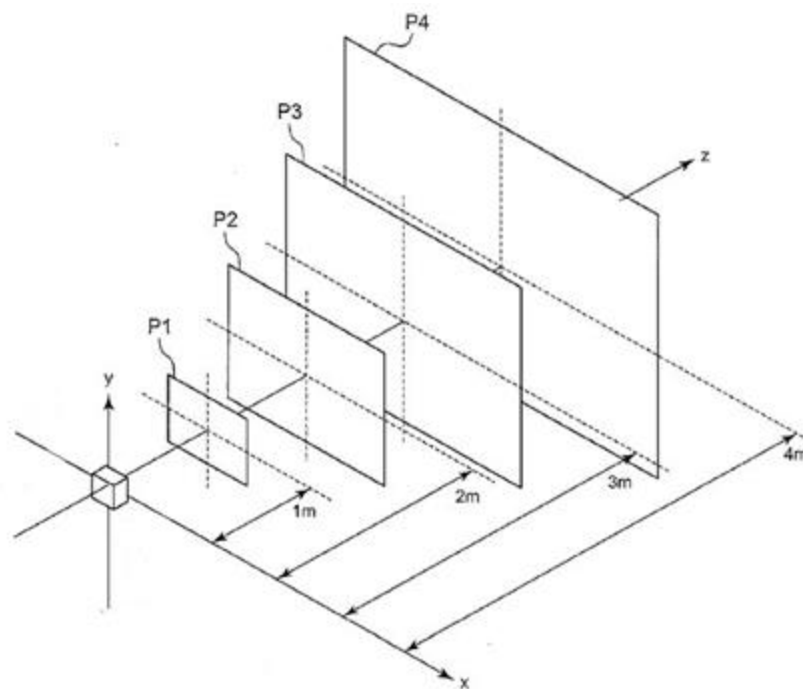
*Figure 6.7: Difference graph between Kinect v2 depth data and the correct position demonstrating the accuracy over distance*

However, the overall magnitude of this movement is only 20mm over the entire depth range and will have a minimal effect on depth measurements between 2m and 2.8m as the data only ranges between 15–23mm. The primary concern demonstrated in this figure is the amount of noise present at any particular depth. This noise was not observed during a similar test while the camera was at the University of Washington medical centre in the USA. As this test was performed at the University of Canterbury in New Zealand (230V ac supply) the Kinect v2 camera used for testing (110v), required an 110v transformer as a power supply. It is unknown if this had any impact on the camera's normal function or whether the camera was potentially damaged during shipping. When the Kinect v2 camera is released to the general public, a 230 volt version will be purchased for use in New Zealand.



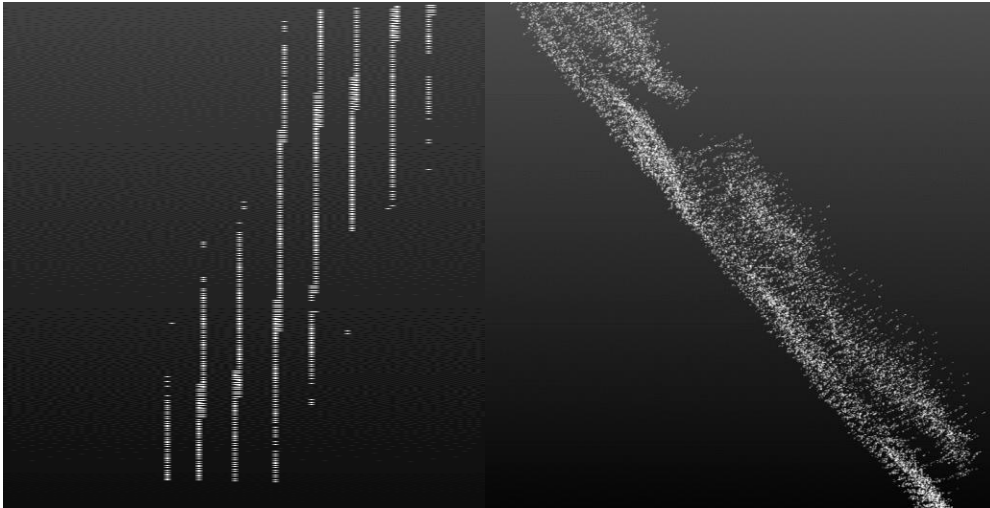
### 6.2.3 Vertical and horizontal resolution

The vertical Y and horizontal X resolution of the Kinect is dependent on the number of pixels and the size of the FOV. With a divergent FOV, which is the case with the Kinect, the greater the distance, the poorer the vertical and horizontal resolution. This is illustrated in Figure 6.8 with a 2 by 2 pixel camera demonstrating the decreasing resolution. Since



*Figure 6.8: Illustration demonstrating that with a divergent FOV the resolution decreases with distance*

the field of view is known and the number of pixels is known, the resolution for the Kinect v1 is 3.1mm per pixel per metre and the Kinect v2 is 2.4mm per pixel per metre. This hardware limitation will result in low accuracy in the horizontal and vertical directions.



*Figure 6.9: Image demonstrating the discreteness observed in the depth data from the Kinect v1 (left) and Kinect v2 (right). This image is viewed from above showing the difference in discreteness observed*

## 6.2.4 Kinect subsampling

The Kinect v1 has a USB 2.0 connection capable of transferring information at a rate of 480 Mbit/s. While the Kinect v2 has a USB 3.0 connection, capable of transferring information at a rate of 5000Mbit/s. This limited bandwidth available to the Kinect v1 forced the output of the Kinect v1 to be subsampled to an 11bit short compared to the 16bit short used in the Kinect v2. This resulted in the data having a discreteness in nature. This effect is shown in Figure 6.9, which is showing 305 frames which have been combined together, obtained from a slanted plane.

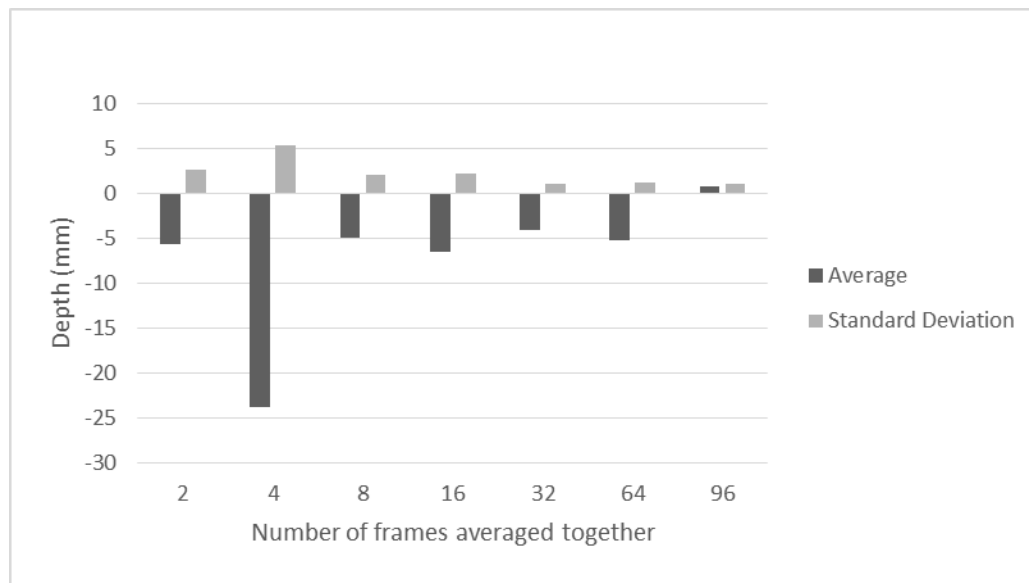
This data is viewed from above or along the Y axis instead of the normal acquired Z axis view. The image should display a slanted point cloud with noise present as demonstrated by the Kinect v2 image (right), however the Kinect v1 image displays a discreet pattern rather than a smooth pattern as a result of the limited bandwidth. This discreteness has been implemented by Microsoft™ at the hardware level so it cannot be reversed.

## 6.3 Software

This section discusses and provides results of tests to the various software components of the system in determining the optimal configuration of the system. This section also discusses both averaging which is performed by the software to reduce noise and the mapping software which ensures the coordinate systems coincide.

### 6.3.1 Kinect noise and averaging

Figure 6.4 demonstrates that the Kinect sensors have an inherent noise, which is Gaussian in nature. This noise can be reduced by using temporal averaging. Therefore a test was devised to determine the amount of averaging required for the system. The camera was positioned a distance of 850mm away from a smooth flat wall and the data collected for one minute for each of the averaging sample rate being tested. The Kinect cameras were allowed the full warmup period before the test was started. Figure 6.10 displays the data obtained from the testing. The averaging method used averaged frames together to produce a noise reduced frame. The number of frames averaged together that were tested, ranged from 2 to 96.



*Figure 6.10 Graph demonstrating the effect of averaging on total noise on the system using the Kinect v1*

The graph clearly shows that increasing the number of frames averaged together decreases the standard deviation and the noise. The only downside to increasing the number of frames averaged together, is this will slow down the response of the system and temporal resolution. It was also decided to use a moving average so that data could be obtained at a high frame rate and still be averaged. Therefore a compromise was decided between reduction of noise and temporal resolution. It was decided to use an averaging of 30, so the system would update once each second, as the camera frame rate is 30Hz. This averaging was used in all future software.

### 6.3.2 Iso-Mapping

The coordinate system generated from the Kinect sensors needed to be transformed from Kinect sensor coordinates to LINAC or couch coordinates in order for movements detected to be meaningful. The Isomapping system developed and discussed in a chapter 5 required testing to determine the accuracy of the transform. As the mapping system is a two-step process both steps were tested. The test required the placement of a Kinect camera in a LINAC bunker at the end of the couch. The Kinect camera was levelled using the room lasers and internal accelerometer. The natural test would have been to rotate the camera around the phantom, however this would be difficult to achieve in practice.

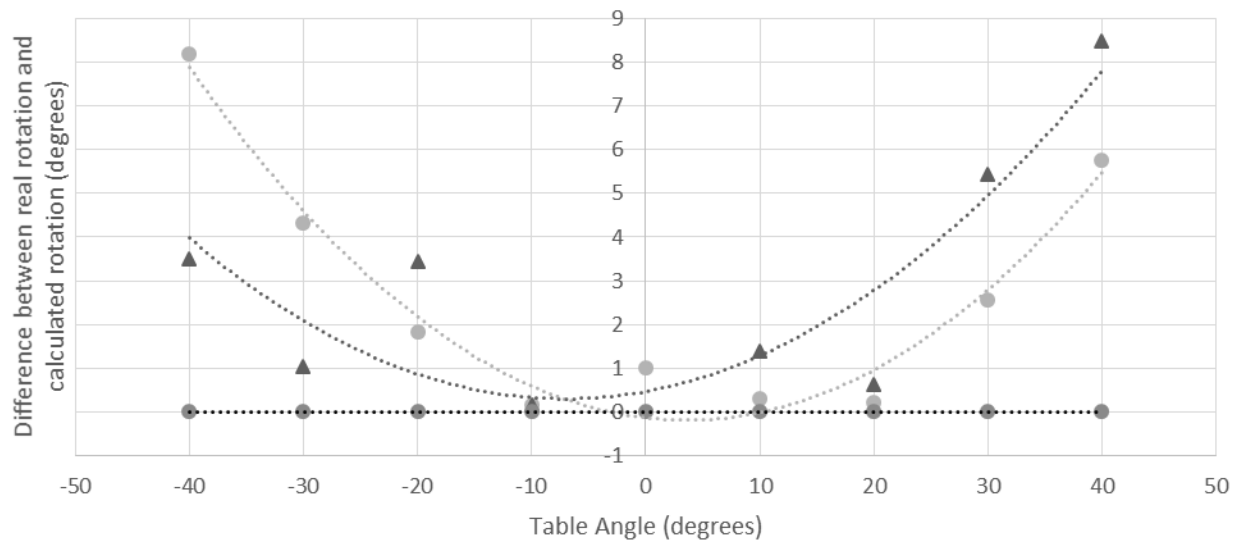


Figure 6.11: Graph displaying the error as a function of rotation. Results from rotation around the Y axis is shown by light colour circles, rotation around the X axis is shown by triangles and the 2<sup>nd</sup> step corrected rotation is shown as dark circles

The same data were instead acquired by keeping the camera fixed in space and rotating the phantom by rotating the couch upon which it was positioned. The initial position as determined by aligning the phantom to the external laser system was defined as the zero rotation position. The couch was then rotated in ten degree intervals from 0 to -40 degrees. The Kinect camera was then rotated about the horizontal axis which aligns the centre of the camera with the isocentre by 90 degrees on the stand and the phantom moved in the same fashion as described for the x axis rotation to simulate a y axis rotation. Three transformations were collected for each point and averaged to provide an improved result. Figure 6.11 demonstrates the rotation error in calculating the real rotation. The x and y rotations show that the rotation error will increase for the first step in the rotation alignment method as the table is rotated to larger angles. While it is assumed this error should be symmetrical a small difference is observed between the negative and positive table rotations, resulting in the fitted curve having a minimum of below the x axis. The Corrected rotations by the second step remove this error. It is unknown how accurate

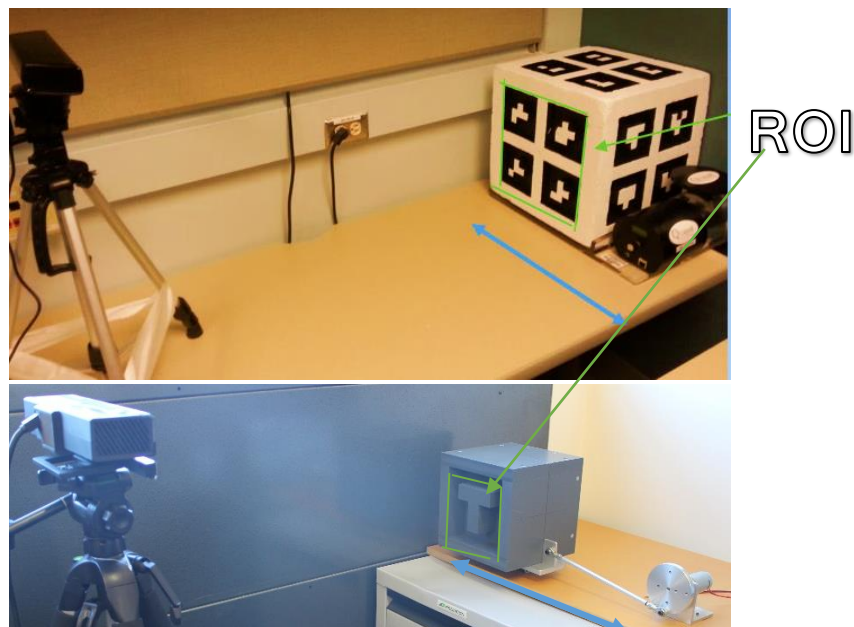
the second step is, as the accuracy of the table rotations was only 1 degree. It is therefore assumed that the entire rotation tracking method is accurate to 1 degree. This Iso-mapping method was applied in all further software.

## 6.4 Tracking software testing

Chapter 4 introduced four tracking methods which were implemented in the software system. Each of these methods were tested in this section to determine their strengths and weaknesses. In order to achieve this an object of repeatable known motion needs to be tracked. This section introduces the motion platforms used in this thesis and then discusses the different tracking methods and results.

### 6.4.1 Motion tracking and motion phantom

Two motion tracking platforms were used to test the tracking components of the system. A motion platform consists of a moveable platform that moves in a single direction parallel to the ground. The displacement of these platforms relative to the equilibrium position moves in a sinusoidal motion with a pre-determined frequency and amplitude. By comparing the known motion (sinusoidal oscillating) of the motion platform, to the measured motion produced by the tracking software, the accuracy of each tracking method can be determined. Figure 6.12 is of the motion platform and phantom used at the University of Washington Medical centre (top) and the University of Canterbury (bottom). As shown in Figure 40, the Camera was placed 850 mm from the front of the phantom, The ROI used is the test is displayed in green, and the movement direction (lateral) is shown in blue.



*Figure 6.12: Images of the Motion breathing platform with a phantom placed on it. The direction of oscillation is shown by the blue arrow and the ROI is shown by the green area. University of Washington setup (top) and University of Canterbury setup (bottom)*



After allowing 40 minutes for the camera to warm up as recommended in section 1.2.1, the software would track the phantom's movement for 1 minute. A similar setup was used at the University of Canterbury with the following differences; the Kinect v2 was used for testing, the Isocentre cube was used as the phantom, and a different motion platform was used as seen in Figure 5.9. The following sections give the results from this testing. Due to time and software restraints the Kinect v2 was not able to be used for all tests so some tests use the Kinect v1.

## 6.4.2 Camshift

Testing of the Camshift method was done at the University of Washington Medical centre as per section 1.3.1 using the Kinect v1 camera. A phantom was placed on the platform which oscillated at a known frequency and amplitude. The Camshift tracking was running in RGB mode for this section. This means the 2D position is determined with the colour camera and this data is then used to find the predicted position with the depth camera. Figure 6.13 shows the data obtained from Kinect tracking as crosses on the figure and the true motion which the platform follows. As the figure shows the Camshift is clearly able to track and follow the phantom. Figure 6.14 shows the residuals from the true motion and the data obtained by the software.

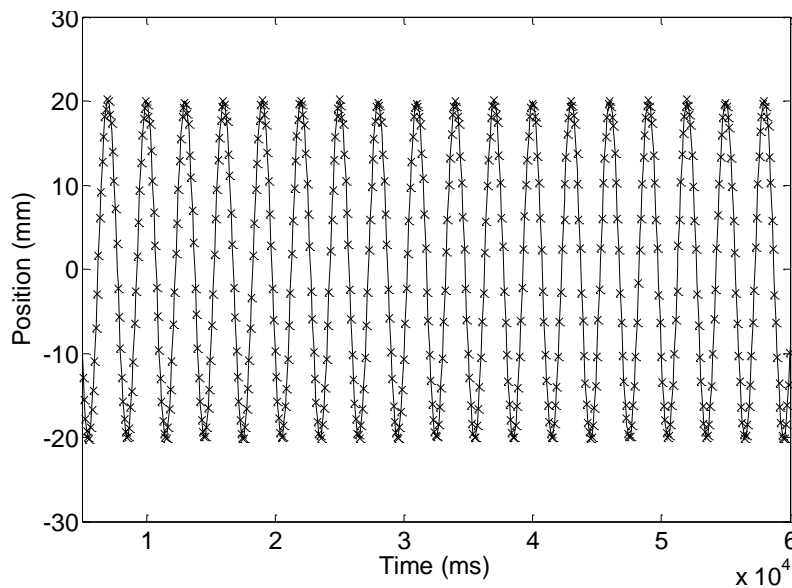


Figure 6.13: Graph showing Camshift tracking data (x) and theoretical position of the motion platform

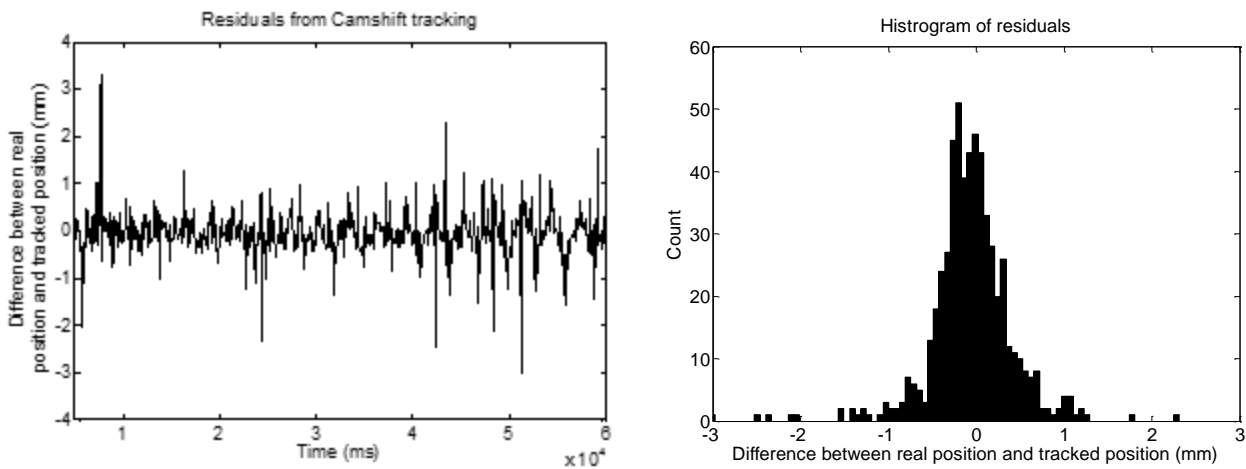


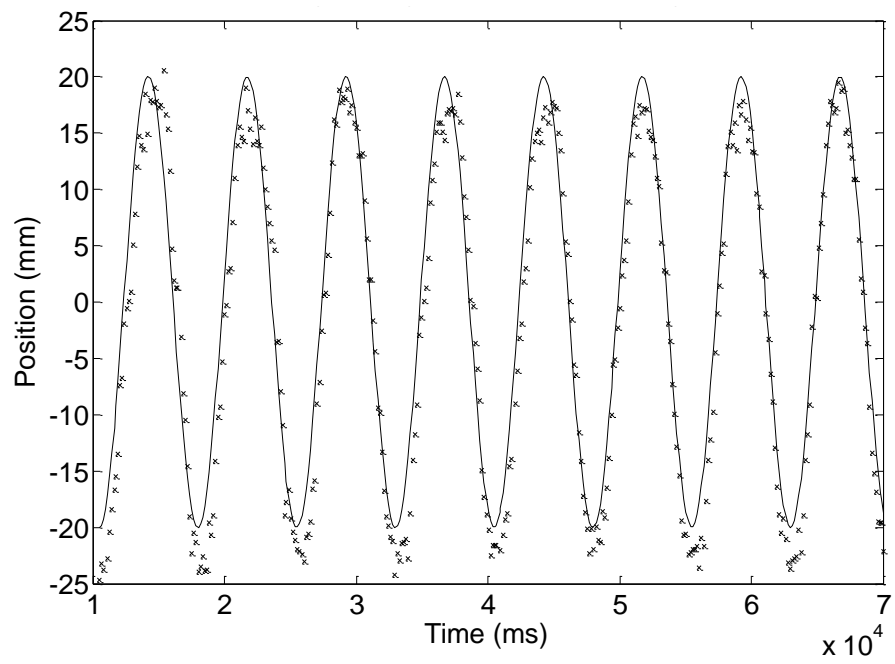
Figure 6.14: Graphs showing the difference between the Camshift tracking data ( $X$ ) and theoretical position of the motion platform

Figure 6.14 shows two images, the image on the left is a time series of the residuals and the image on the right is a histogram plot of the residuals. These data have a standard deviation of 0.5mm and a mean of  $-0.1\text{mm}$ . This clearly demonstrates that in an office with bright lighting the Camshift method can meet the requirements. However after testing in a clinical environment in this situation a LINAC bunker with poorer lighting conditions than a standard office, the results showed that the tracking software would often lose the target and move away from the phantom and track a different object in the FOV. This occurs due to the decrease in the Value in the HSV images, which has a similar effect to decreasing the contrast of different colours and results in different colour objects being indistinguishable. A second method using the Camshift method was attempted. This involved using only the depth images for tracking (the colour image data were not used). This was achieved by assigning different depth values a particular colour. However this method had issues tracking an object in the camera depth direction as the object changing in depth and would appear to be changing colour. From these results it was determined that the Camshift method is not suitable for use in a LINAC bunker without increased lighting.

### 6.4.3 Kinect Fusion and ICP

Testing of the Kinect Fusion method was done at the University of Washington Medical centre. This was again achieved with a breathing motion platform for the Kinect v1 camera. The standard set was used using the Kinect v1. A phantom was placed on the platform which oscillated at a known frequency and amplitude. The Kinect Fusion tracking uses ICP for the tracking components and has the reconstruction components disabled.

**Error! Not a valid bookmark self-reference.** displays the sinusoidal curve from the motion platform and the Kinect fusion tracking data points. As the figure demonstrates, the Kinect fusion tracking mostly keeps up with the platform but fails at the peaks. This effect is clearly shown in Figure 6.16 where a large range of residuals is observed. The residuals have a mean value of  $-1.7\text{mm}$  and a standard deviation of  $3.0\text{mm}$ . This clearly demonstrates that the Kinect fusion is not accurate enough for the system. This is most likely a result of only a small FOV with very little variation in the scene present. The Kinect Fusion tracking method uses only the depth data obtained from the Kinect and is therefore not influenced by lighting conditions. The Kinect v2 is not yet able to run the Kinect Fusion method as it has yet to be released in the SDK by Microsoft™.



*Figure 6.15: Graph showing Kinect Fusion tracking data (x) and theoretical position of the motion platform*

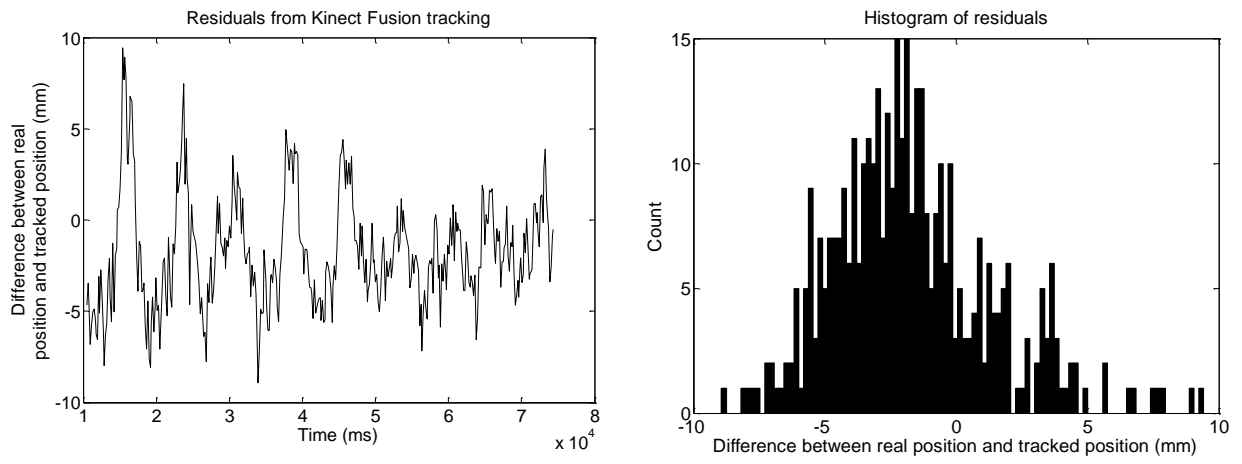


Figure 6.16: Graphs showing the difference between the Kinect Fusion tracking data and theoretical position of the motion platform

#### 6.4.4 Least squares

Testing of the Least squares method was done at the University of Canterbury. This was done with a motion platform and the Kinect v2 camera. This platform was set up to move in a sinusoid motion along the x direction in the Kinect camera's coordinate geometry. The Kinect camera was placed at a distance of 750mm away from the motion platform and levelled. The isocentre phantom was placed on the platform which oscillated at a known frequency and amplitude. Figure 6.17 displays the sinusoidal curve from the motion platform and the least squares tracking data points. As the figure demonstrates the least squares tracking method tracks well, however the peak position of the amplitude is consistently too large. This is further observed in plot of residuals shown in Figure 6.18 where the peaks and the troughs give a sinusoidal appearance.

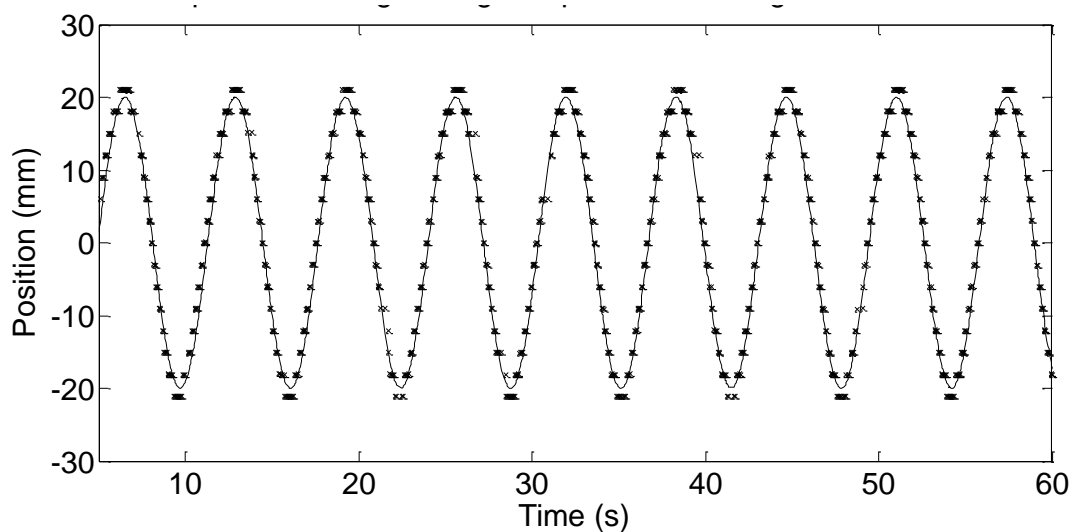
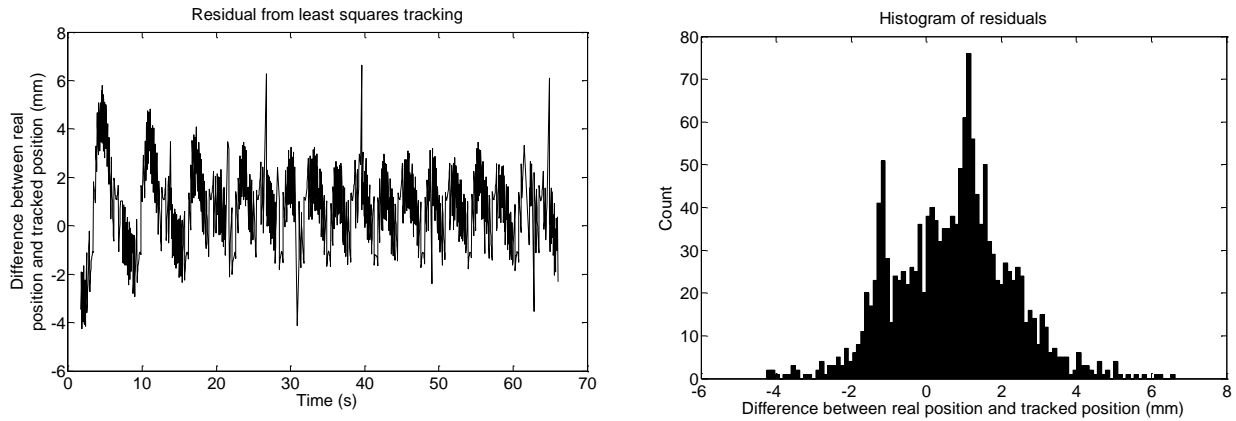


Figure 6.17: Graph showing Least squares tracking data (x) and theoretical position of the motion platform



*Figure 6.18: Graphs showing the difference between the Least squares tracking data and theoretical position of the motion platform*

The least squares residuals have a mean value of 0.7mm and a standard deviation of 1.5mm. While this method is not as good as the Camshift RBG method, it has the advantage that it is lighting independent as it only uses the depth camera (infrared). This method is also twice as accurate as the Kinect fusion method and requires less processing time.

#### 6.4.5 Surf

Testing of the SURF method was done at the University of Canterbury. This was done with a motion platform and the Kinect v2 camera setup as stated in section 6.4. Figure 6.19 displays the true motion of the platform and the acquired least squares tracking data points.



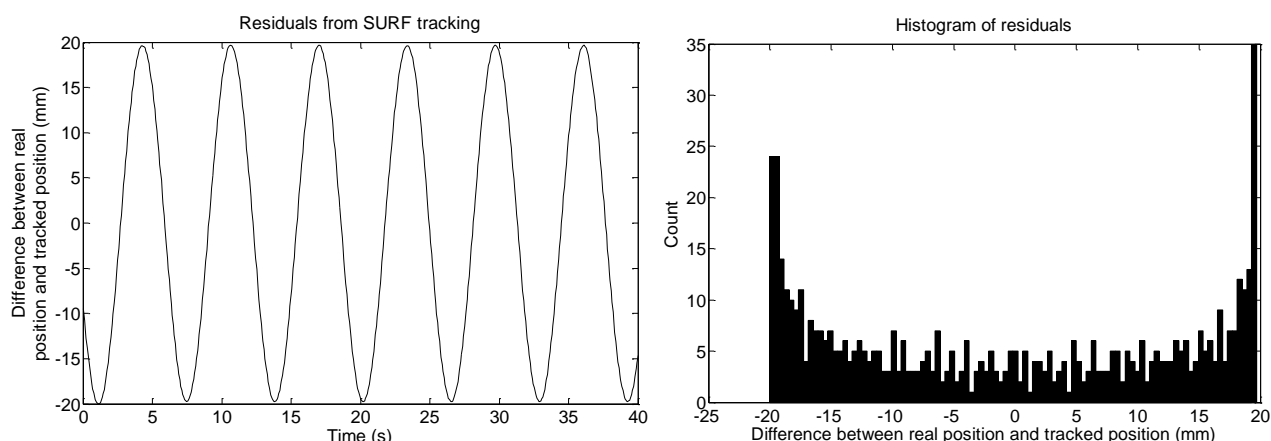


Figure 6.20: Graphs showing the difference between the Surf tracking data and theoretical position of the motion platform

As the figure demonstrates the SURF tracking was unable to track the target. A small sinusoid is observed but the SURF tracking system could not track.

This is most likely due to the lack of key points in the tracked phantom. Figure 6.20 shows the residuals from the SURF sinusoid. The residuals have a mean value of  $-0.9\text{mm}$  and a standard deviation of  $14.1\text{mm}$ . The data demonstrates that this tracking method cannot be used in this situation.

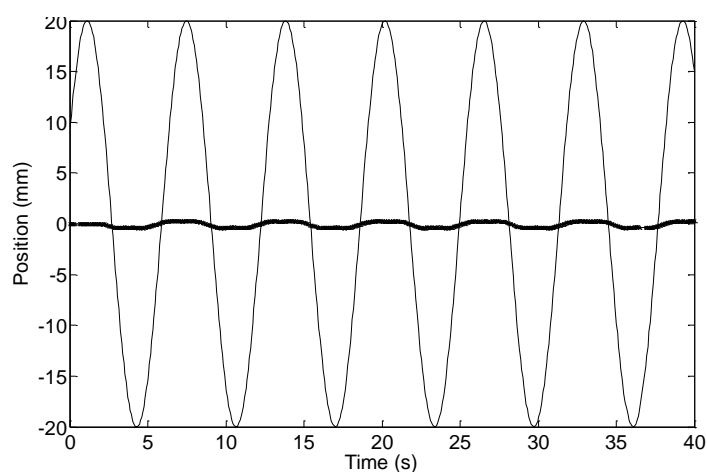


Figure 6.19: Graph showing Surf tracking data ( $x$ ) and theoretical position of the motion platform

## 6.5 Summary

The aim of this chapter was to discuss the various options which were available via the SDK and could be used in the development of a tracking system utilising the Kinect family of 3D cameras with the aim of determining an optimal combination. The first section of this chapter discussed the hardware available for use, the Kinect v1 and the Kinect v2. A summary of the specifications for the two sensors is provided in the following table

	Resolution RGB	Resolution Depth	Depth noise (SD of 1m <sup>2</sup> square at 1.5m distance )	Depth technology	Warm- up time
Kinect v1	640x480	640x480	2.4mm	Triangulation	60min
Kinect v2	1920x1080	512x424	0.6mm	Time of flight	60min

*Table 1: Kinect hardware comparison*

From the above table it can be observed that the Kinect v2 has a significantly lower depth noise. As this is of primary importance the optimised system will consist of the Kinect v2 and thus only this camera was considered for testing the tracking routines. The results of which are shown in the following table:

	Standard deviation ( mm )	Relative speed	Lighting dependant	Accuracy	Ability to deal with deformations	Degrees of freedom
Camshift	0.5	Fast	Yes	Very High	High	4
SURF	14.1	Medium	No	Very Low	Moderate	4

Kinect Fusion	3.0	Slow	No	Medium	Low	6
Least squares	1.5	Fast	No	High	Moderate	3

*Table 2: Tracking methods comparison*

Table 2 shows the key differences between the different tracking methods as applied with the Kinect v2 camera. The Camshift method performed the best overall. However, this method was found to be sensitive the ambient lighting levels and this limits its use in LINAC bunkers and consequently cannot be used for this system. The Least squares method is the next most accurate method tested and as this method is not sensitive to the lighting levels this tracking method will be used when testing the entire system in a clinical environment. A limitation of this method is that it only tracks with 3D of freedom and it has been assumed that the lack of rotational tracking will not negatively impact the system. This seems reasonable as patients should not rotate from their initial position during treatment. Further improvements in the Kinect fusion method by Microsoft™ could result in this implementation of the ICP being used instead of the Least squares as it has six degrees of freedom however at time of writing ICP is not suitable. In addition, this method is currently not unavailable on the Kinect v2.

The next chapter will use the results determined from this section to test the entire system in a clinical environment using the now determined optimal tracking configuration consisting of the Kinect v2 camera and the least squares method of tracking.



# Chapter 7: Clinical Testing

---

This chapter introduces results from testing the Kinect cameras and dedicated software in a LINAC bunker at both the University of Washington Medical Centre, Seattle, USA and St Georges Hospital, Christchurch, New Zealand. This chapter is divided into four sections: clinical setup, position testing, volunteer testing, and 4D-CT phantom testing. The clinical setup section will discuss the process and methodology of setting up the system in a clinical environment. The section on position testing will cover testing the system's ability to track a fixed phantom on the LINAC couch in the lateral, longitudinal and vertical directions. Finally the volunteer testing section will discuss testing of the system in a clinical environment to simulate real situations which could occur during treatment.

## 7.1 Clinical Setup

During clinical setup the Kinect v2 sensor was attached to a tripod which is ideally positioned at the end of the LINAC couch. When this was not possible due to, for example, the LINAC bunker design, the tripod was placed as close to the end of the couch as possible. The tripod's height was adjusted so the Kinect sensor was able to view the isocentre from above. This was usually achieved by placing the camera at the maximum height of the tripod which is around 170cm. The Kinect sensor was connected to the laptop which for these tests is able to be placed conveniently in the room as there is no intention to beam on. The software and the Kinect v2 sensor were then started to ensure the correct warm up period. While it was desirable to provide the full 60 minute warmup period for the sensor, in some testing situations there was insufficient time available due to limitations of interrupting a clinical environment. Therefore only a short warmup period of 15 minutes was achieved in this situation.



*Figure 7.1: Image of camera placement and setup at the proton therapy centre, Seattle*

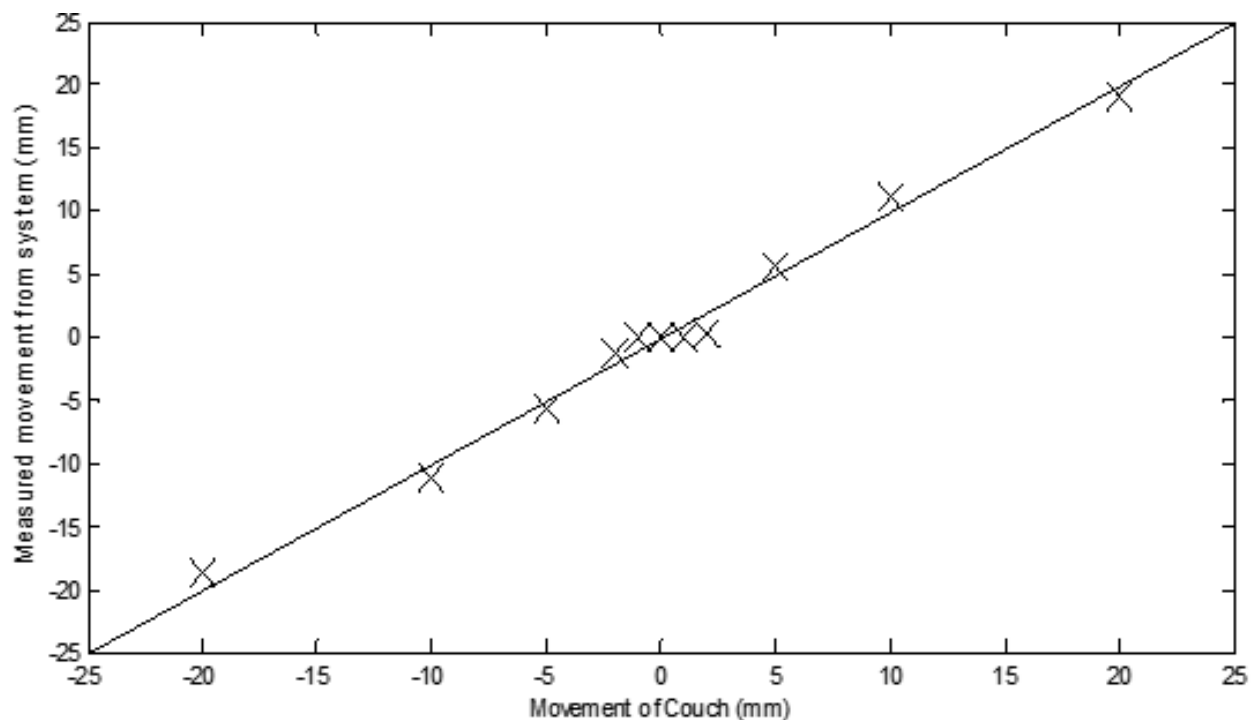
Figure 7.1 shows the placement of the UW Isocentre cube at the LINAC Isocentre, the Kinect v2 sensor placed at the end of the table, and a laptop connected nearby. The Kinect v2 sensor is ideally placed as close as possible to the isocentre as shown in Figure 49. However, limits on tripod heights over the table often prevented this. The next step in the setup was mapping the sensor coordinates to LINAC coordinates via the process as described in section 5.5. This is to ensure all data outputted from the Kinect sensor is in directly applicable coordinates. The process has been discussed in the previous chapter and uses an Isocentre cube placed at the LINAC isocentre. The rough alignment is performed in software by the user selecting the four corners of the cube. A fine alignment is then completed using the ICP method in the Cloud compare software. Once this coordinate mapping is complete, the mapping is saved and the Isocentre cube can be

removed but the Kinect camera must remain in its current position. This concludes the setup method and the software is now ready to track and monitor patient or phantom positions.



## 7.2 Position testing

This section of the thesis will introduce and discuss the testing of position measurements. For a patient positioning monitoring system to be effective it must be able to detect and measure real patient movements in the lateral, longitudinal and vertical directions. Therefore the following test was devised to assess how accurately the Kinect sensor recorded motion in these three orthogonal directions. A stationary, rigid phantom was placed on a LINAC couch, and incrementally moved by  $\pm 1, 2, 5, 10$  and  $20$  mm in each of the three directions: independently. The size of the increments was determined to assess both large and small movements. This testing was performed using the Kinect v2 sensor and the



*Figure 7.2: Testing of the system with movements in the lateral direction*

least squares tracking method.

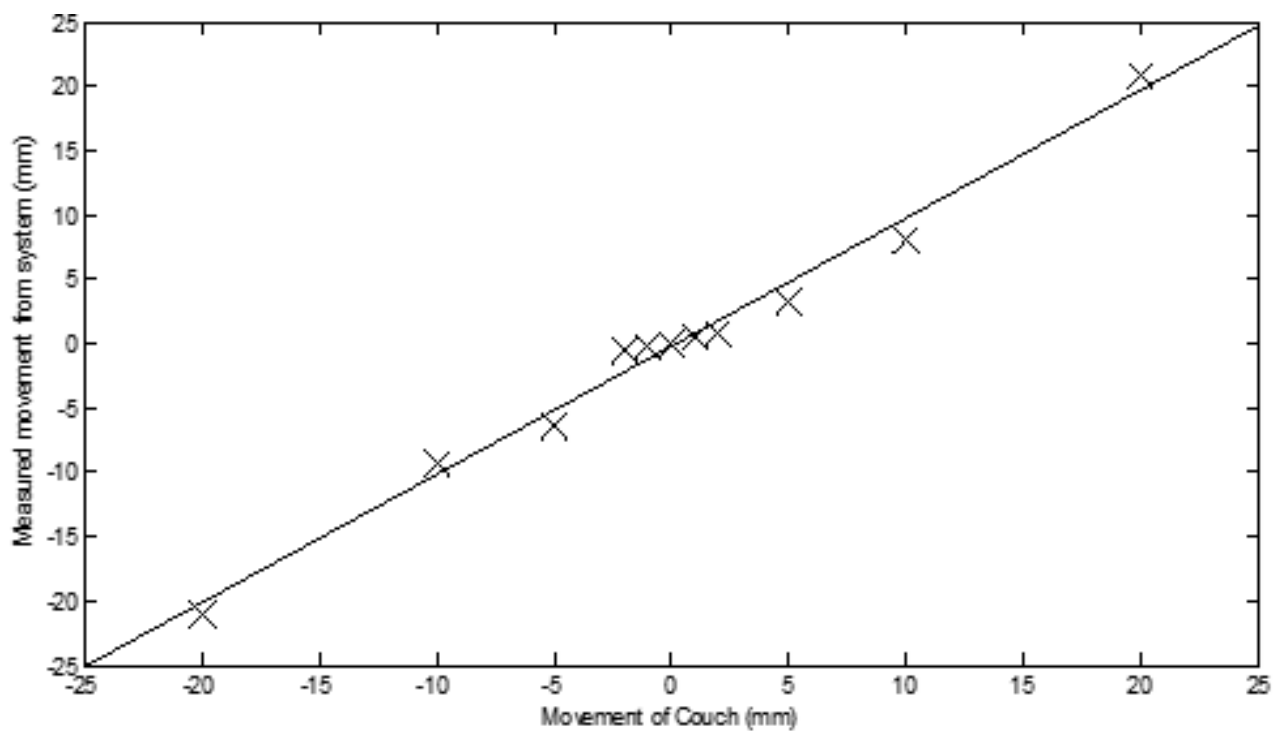


Figure 7.3: Testing of the system with movements in the vertical direction

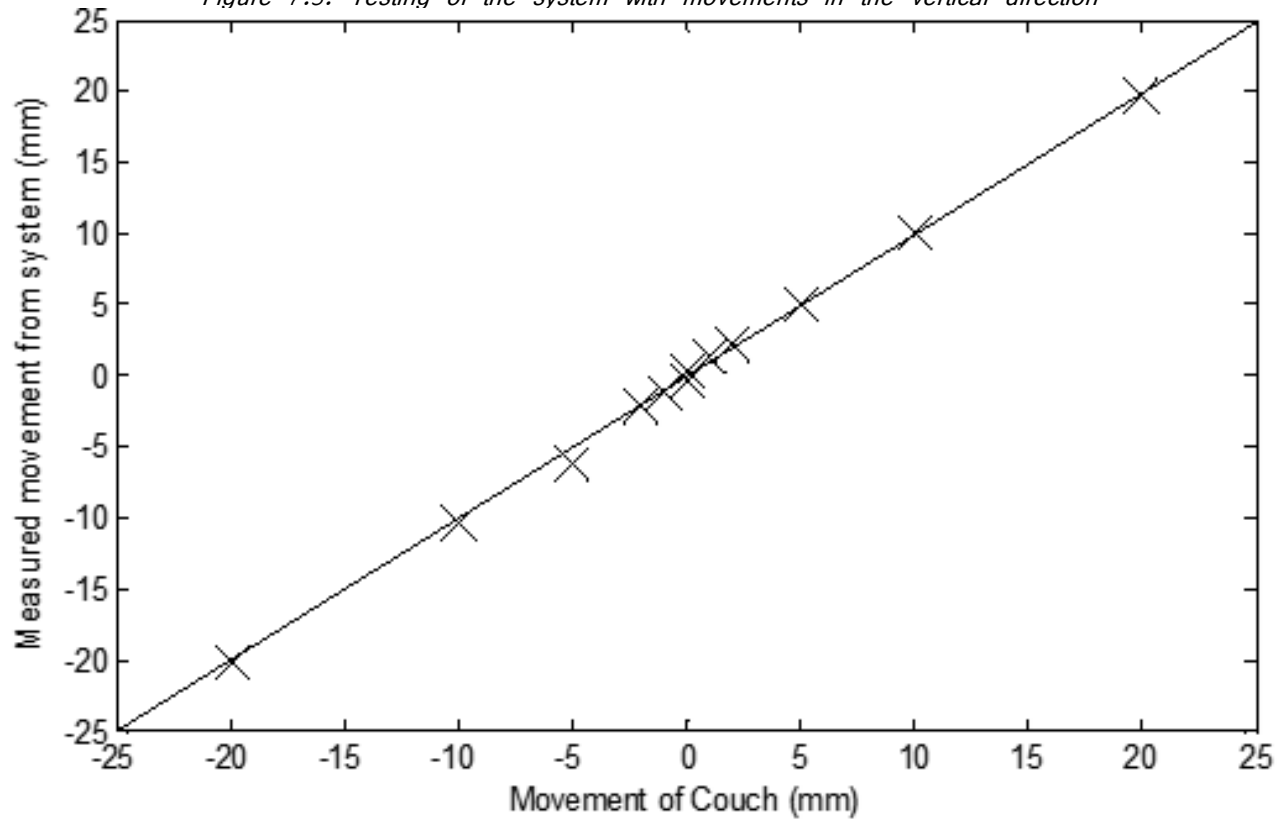


Figure 7.4: Testing of the system with movements in the longitudinal direction

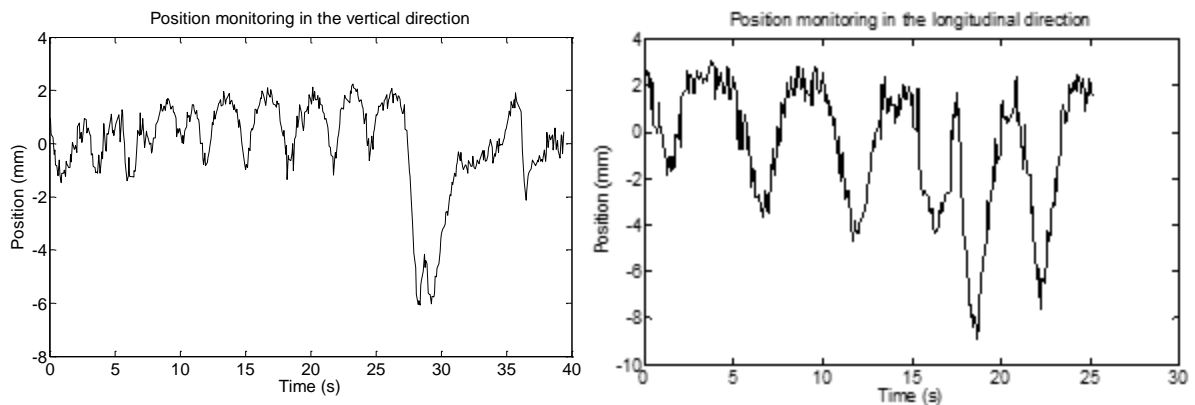
Figure 7.2, Figure 7.4, and Figure 7.3 results from the lateral, longitudinal and vertical movements respectively. The longitudinal movements are in the z direction for the Kinect sensor. Therefore this corresponds to the depth component of the system. As the results demonstrate, the system was able to track in the longitudinal direction with high accuracy. The lateral and vertical movement results are very similar, and these correspond to the x and y directions from the Kinect v2 sensor. These results demonstrate that large movements can be accurately detected by the system. However, the low spatial resolution of the sensor limits the system's ability to detect small movements, resulting in a flat region around zero. The five millimetre movement is clearly visible in all directions, as are all movements greater than this. The system is therefore capable of tracking in all 3 directions and can detect movements greater than 5mm.

## 7.3 Volunteer testing

For this section of the thesis a volunteer was asked to lie on the treatment couch in the supine position. The volunteer was then monitored and asked to perform various motions which could realistically occur during treatment. The system monitored their position and determined if any change in position could be detected. For this test it was decided that any patient movement greater than 10mm in any or all of the three orthogonal directions would have a negative impact on the theatrical treatment. In many of the situations there was little or no movement in the lateral direction. Therefore these graphs are not shown. All data shown have only a two frame average applied. This is to ensure fast movements are not missed. The ROI selected to track was an upper section of the volunteer's chest, representing a lung tumour being treated. This ROI was chosen as it is a highly mobile region and can be difficult to treat. Tests were devised starting with very small movements moving onto larger movements to determine whether there was a threshold of minimal movements able to be detected.

### 7.3.1 Baseline

For this test the volunteer was asked to remain as still as possible to establish a baseline for their breathing cycle. As Figure 7.5 illustrates, the system is capable of detecting the small breathing motion of the volunteer. At time 28 seconds the volunteers breathing was observed by the system to change so this test needed to be repeated to determine whether this was an error in the system or the volunteer did change their breathing. No movement was observed in the lateral direction, as was expected for this test. Of note, the volunteer

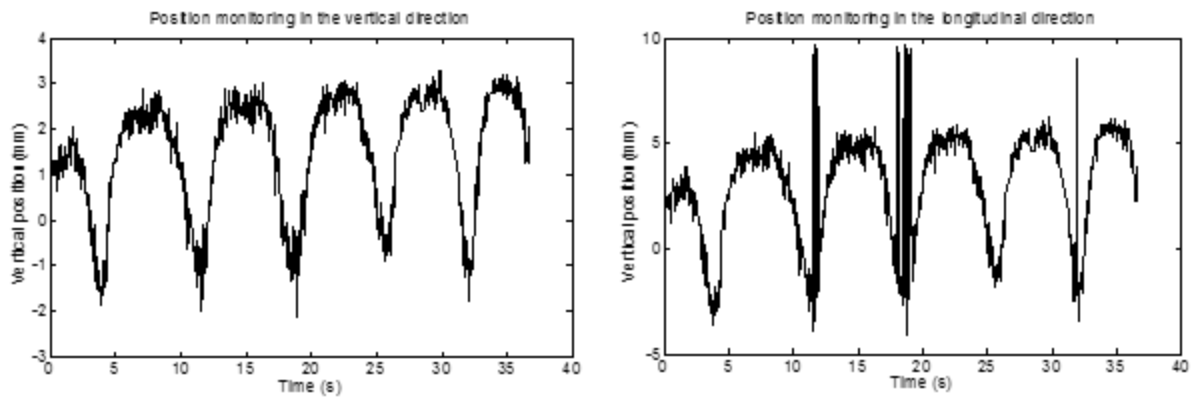


*Figure 7.5: Base line movements of the volunteer over 25 seconds.*

did not move outside the tolerance during this baseline test.

### 7.3.2 Normal breathing

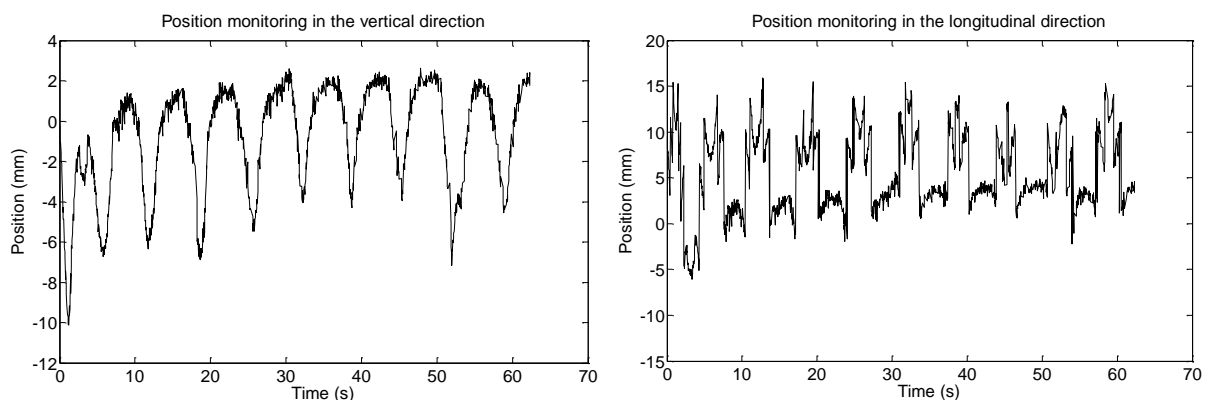
For this test the volunteer was asked to remain still and just breathe normally to determine if the baseline was accurate. This test should represent what occurs during most treatments. The results were very similar to the baseline test as expected except no change of breathing is observed ruling out our system as the cause of the change. Figure 7.6 displays the data obtained during this test. Some noise is present in the system as seen by the vertical bars in the longitudinal direction graph at time 12, 18, and 32 seconds. This noise however, does not go beyond the set threshold of 10mm. The cause of the noise in this situation is currently unknown and was only observed in this test.



*Figure 7.6: Normal breathing of the volunteer over 37 seconds*

### 7.3.3 Heavy breathing

For this test the volunteer was asked to increase their breathing to simulate either a patient that is a heavy breather or a patient under stress. Both of which can be quite common in daily treatments. Figure 7.7 shows the data obtained from the heavy breathing test. It can be observed that during the breathing cycle, the volunteer was outside the present tolerance in the longitudinal direction at the peak of the breathing cycle. If this went on undetected, it could have a negative impact on the patient's therapeutic treatment. These results demonstrate that the system can detect these movements during a treatment, and could inform the therapist, who in turn would take appropriate action as dictated in local protocols.

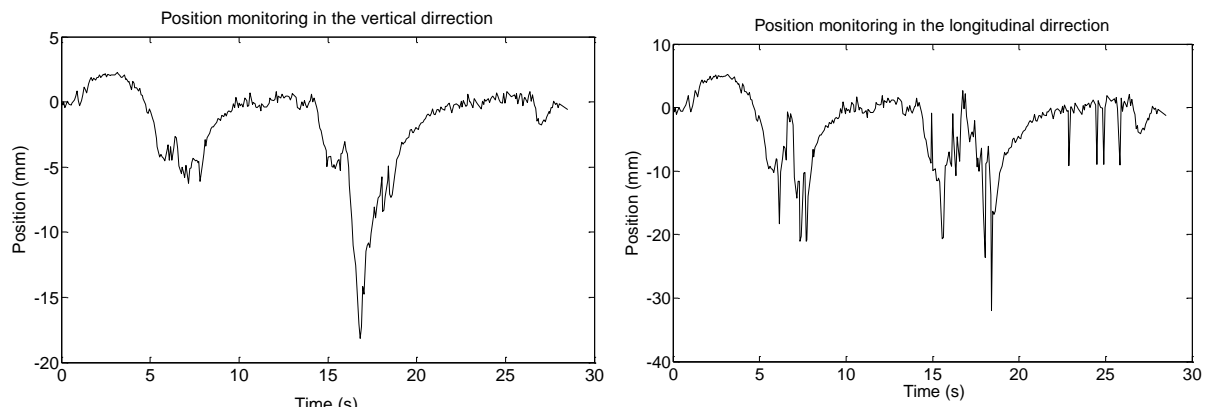


*Figure 7.7: Heavy breathing of the volunteer over 65 seconds*



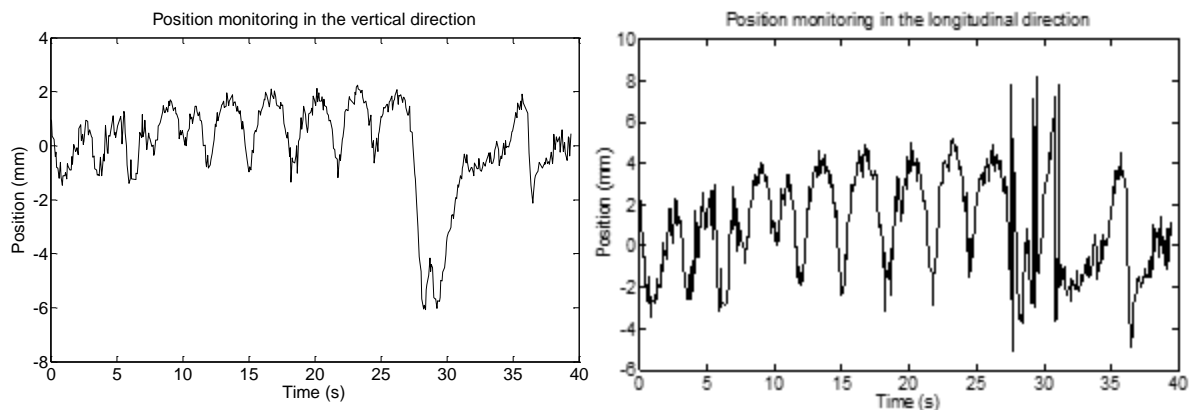
### 7.3.4 Coughing

Due to throat irritation patients may cough during treatment and this could impact negatively on the patient's treatment. Coughing is generally associated with extensive movement and can be short or long in duration. Ideally such a system should be able to identify this form of motion and determine if the motion is beyond the designated tolerance. For this test the volunteer was asked to cough. Figure 7.8 displays data obtained from this coughing test. The volunteer's coughing occurred during 5 to 10s and 15 to 20 seconds. This is quite visible in Figure 7.8, and shows that during this time the volunteer has moved well



*Figure 7.8: Coughing testing of the volunteer, volunteer was coughing between time 5–10s and time 15–20s*

outside the defined tolerance of 10mm. If this continued to happen over the course of a treatment, it could have a large negative effect on the delivered dose distribution and therefore patient outcome.



*Figure 7.9: Volunteer looking around of the volunteer for a duration of 40 seconds*

### 7.3.5 Looking around

This test was to determine if a patient moving their head and looking around while attempting to keep their chest as still as possible, would be detected by the system.

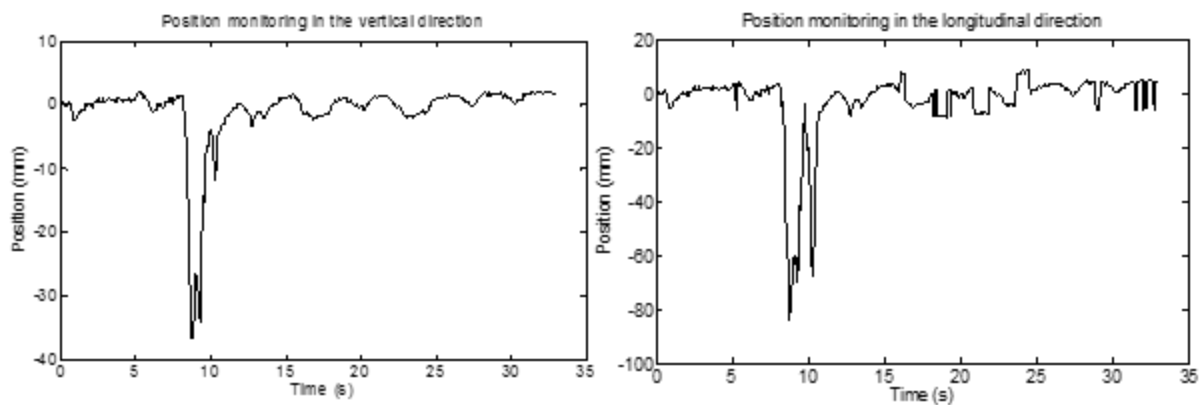
Figure 7.9 displays the results from the test, and it can be seen that the volunteer did not move outside of the tolerance. However the longitudinal motion has increased.

### 7.3.6 Moving their backside

As patients are often lying on the treatment couch for long periods of time, they can become uncomfortable. Therefore, they will often readjust their backside. This can have a large impact on the patient's position during treatment. The volunteer was asked to move around and this is observed in Figure 7.10 at the 7 second mark. During this time the volunteer is well outside the tolerance. However, after the volunteer has moved they have returned back to their initial position.

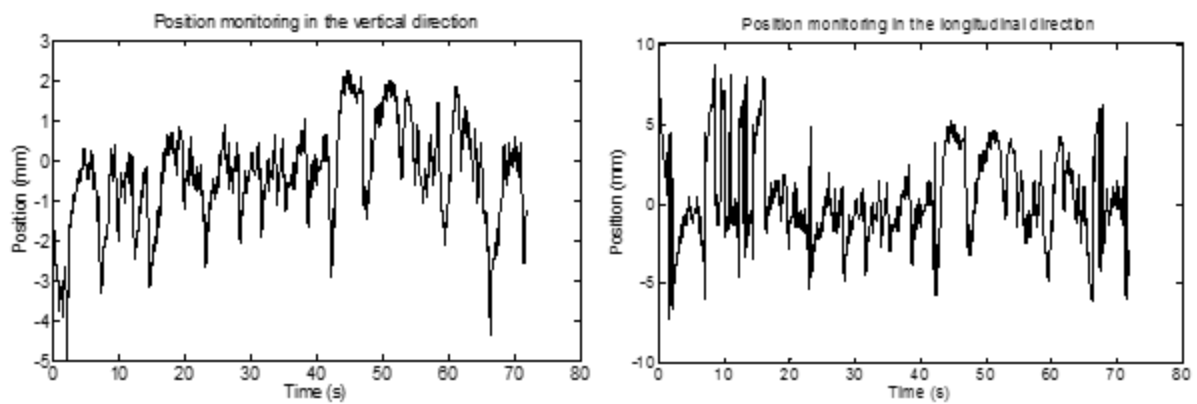
### 7.3.7 Talking

During this test the volunteer was asked to talk throughout the test. This was in order to ascertain whether the system was capable of detecting this type of motion and whether it has a large impact on the volunteer's position throughout treatment. Figure 7.11 displays the data obtained from this test. It can be observed that throughout the test the patient position never went outside of the tolerance. However, it is clear that the volunteer's motion in the longitudinal direction has increased.



*Figure 7.10: Volunteer readjusting backside at time 8 seconds then returning to original position*

### 7.3.8 Arm movements



*Figure 7.11: Volunteer talking for a time of 75 seconds*

For this section of testing the volunteer was asked to move their arms about. This was to determine if the system was capable of detecting this type of motion and measure if this has an impact on the theoretical treatment. Figure 7.12 displays the data obtained from this test. It can be observed that during the test the volunteer did move outside the 10mm tolerance for a short period of time. The system is clearly able to detect this impact on the chest from the motion of the volunteers arm movements and it would have a clear impact on treatment.

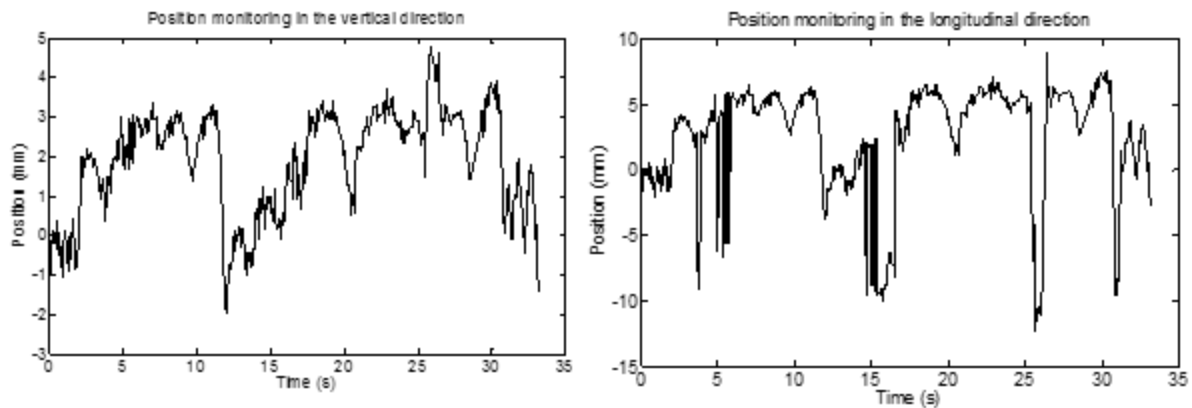


Figure 7.12: Volunteer moving their arms for 33 seconds

### 7.3.9 Summary

Table 3 displays the main results from testing with a volunteer in near identical conditions as per treatments, only difference was no radiation beam was activated during the test. As of yet the effects of scattered radiation has not been tested on the sensor as we do not currently have a replacement if damage was to occur. From this table it can be determined that in all cases the system was able to detect the motion generated. In four of the test situations the volunteer's movement was beyond the 10mm tolerance and this was observed by the system identifying this had occurred by displaying a red warning. The system is also clearly able to track the volunteer's breathing throughout testing.

	Detectable	Beyond tolerance	Comparison to baseline	Maximum shift from initial position (direction)
Normal breathing	Yes	No	Very similar	8mm (long)
Heavy breathing	Yes	Yes	Significantly increased motion	15mm (long)
Coughing	Yes	Yes	Large motion during coughing	30mm (long)

Looking around	Yes	No	Increased motion	8mm (long)
Moving bottom	Yes	Yes	Massive disruption followed by return to baseline	80mm (long)
Talking	Yes	No	Increased motion	8mm (long)
Moving Arms	Yes	Yes	Largely increased motion	12mm (long)

*Table 3: Volunteer testing results*

## Chapter 8: Discussion and Conclusion

---

This chapter concludes the thesis by discussing results from hardware testing, the system developed in this project, the results from testing, the resulting hardware limitations, and concluding with possible future work.

## 8.1 Hardware

Testing of the system determined that the Kinect v2 was the better of the two sensors assessed for monitoring patient movement during external beam radiation therapy. This is primarily due to the significantly lower image noise when compared to the Kinect v1 and was demonstrated in section 6.2.1 when long term testing of the Kinect v1 demonstrated a standard deviation of 2.4mm, compared to the Kinect v2 of only 0.6mm. Both sensors require a warm up period and this was determined to be 40 minutes. After this period both sensors were observed to be very stable. If used in a clinical environment the sensors could be turned on in the morning before treatment and allowed the necessary warm up time. After this time the sensors are capable of running the entire length of a clinical workday. The accuracy of the Kinect v2 was observed to vary by over 20mm over 2100mm. However, as the sensor will only be operating over a small range, the impact of this inaccuracy should have a minimal effect on measurements of patient motion. The primary hardware limitation in the Kinect v2 sensor was determined to be the low horizontal and vertical resolution. This resulted from the large FOV in the sensor. The large FOV was unneeded, as only a small  $1\text{m}^2$  area at a distance of 3 metres was needed to be observed. Thus all pixels which were outside this area were unused. The Kinect v2 has also removed the discreteness in the depth data which was observed in the Kinect v1. There are likely to be many reasons for this improvement – one of which is likely due to the higher bandwidth available from the new USB 3.0 interface.

## 8.2 Software

Testing the software components of the system determined that individual frames are inherently noisy and to reduce this noise a large number of frames needed to be averaged together. However, applying this smoothing technique also decreased the response time of the system. As such a compromise was required and the number of frames chosen to be averaged together was set to 30. This enabled the data rate of the mean position to be 1Hz and this was deemed acceptable. The coordinate transformation technique developed for this project was a two-step system. The first step was able to perform an initial alignment which is accurate to within 8 degrees. The second step reduces this error down to less than 1 degree of rotational inaccuracy. This allows the system to operate in LINAC coordinates which is for a requirement for therapists.

### 8.2.1 Tracking software

The investigation into determining the optimal tracking solution for this project concluded that the least squares method provides the most accurate tracking in the clinical environment due to lighting restrictions. This method however only provides translational tracking and assumes that there is no rotational motion of the patient. The Camshift method would be preferred as a tracking method due to its inclusion of the rotational component of tracking and its increased accuracy, however the poor lighting in the LINAC bunker has meant this method as currently implemented within the SDK is not useable in the clinical environment. The below table from Chapter 6 summarises the results from the tracking section.



	Standard deviation (mm)	Relative speed	Lighting dependant	Accuracy	Ability to deal with deformations	Degrees of freedom
Camshift	0.5	Fast	Yes	Very High	High	4
SURF	14.1	Medium	No	Very Low	Moderate	4
Kinect Fusion	3.0	Slow	No	Medium	Low	6
Least squares	1.5	Fast	No	High	Moderate	3

*Table 4: Tracking methods comparison*

### 8.2.2 Position and Clinical testing

Results from the position testing performed by moving a LINAC couch demonstrated that the depth accuracy of the system meets our requirement of being able to detect 1mm movements. However, the accuracy of horizontal and vertical movements as detected by the system was limited by the sensors low resolution resulting from the large FOV. This was primarily observed in the system as not being able to detect shifts smaller than 5mm. It can be concluded that the overall accuracy of the system in its present state is 5mm. This accuracy means that the system could be used in 3DCRT and IRMT as these treatments have around 1cm tolerance. However the system is not accurate enough for SBRT in its current state.

The below table from Chapter 7 summarises the tests which were carried out with a volunteer to determine if the system could detect various types of movements and whether these movements would have an impact on the patient's treatment. As the table indicates all tests were detectable by the system. When the volunteer moved beyond the 10mm tolerance in four of the tests, this movement was detected and reported. This test demonstrates the system was clearly capable of detecting small movements of a patient including monitoring the patients breathing cycle.

	Detectable	Beyond tolerance	Comparison to baseline	Maximum shift from initial position (direction)
Normal breathing	Yes	No	Very similar	8mm (long)
Heavy breathing	Yes	Yes	Significantly increased motion	15mm (long)
Coughing	Yes	Yes	Large motion during coughing	30mm (long)
Looking around	Yes	No	Increased motion	8mm (long)
Moving bottom	Yes	Yes	Massive disruption followed by return to baseline	80mm (long)
Talking	Yes	No	Increased motion	8mm (long)
Moving Arms	Yes	Yes	Largely increased motion	12mm (long)

*Table 5: Volunteer testing results*

In conclusion the system has demonstrated that with low cost consumer hardware, a tracking system capable of measuring and monitoring patient motion for 3DCRT and IRMT is achievable with a Microsoft™ Kinect v2. Further work to improve accuracy is required so that the system is able to detect very small movements in order for use in SBRT.

## 8.3 Future work

As the primary limitation of this system is the hardware capabilities of the Microsoft™ Kinect v2. This limitation can be addressed by either using a new more accurate sensor or by improving the Kinect v2. The primary reasons the Kinect v1 and v2 were chosen was their low cost and widespread availability, changing to a different sensor would most likely be costly thus limiting the systems availability. Improvement of the Kinect v2 could be achieved by reducing the sensors FOV with optics. A few simple tests at the University of Canterbury demonstrated that this could be achieved with a lens placed in front of the infrared receiver magnifying the sensor. No lensing is required for the emitter, the primary drawback with this method is the whole camera would need to be recalibrated which would not be a simple task.

Further applications for the system developed are; use in tracking breathing for either gating in a LINAC bunker, or development of 4DCTs during treatment planning. As the system can track a whole surface it provides benefits over traditional breathing tracking method which only track a single point. The system demonstrated that it was clearly capable of tracking the volunteer chests movements for both shallow and heavy breathing, applying the system for this purpose would be easy to implement and provide a low cost replacement to current systems with improved capabilities.

Supplementary work could also include importing the surface from the patient's planning CT to be used as a reference surface. This reference surface would be compared to the surface data obtained from the system to produce a more accurate alignment. This would also allow the system to be used to assist the initial setup of the patient in the predefined position similar to the system Talbot developed. Developing this idea further to include the external surface from the 4DCT scan, could provide a dynamic comparison.

This information could in turn be used to produce a combined gating and tracking system, informing the therapists when the patient is in the ideal position both spatially and temporally (in the ideal breathing phase). Using this dynamic surface approach dynamic DRRs could be produced in real time to verify the tumour is in the predicted position using megavoltage beam eye view imaging.

# Bibliography

---

- [1] Ministry of Health, "Cancer statistics," Cancer Society of New Zealand, 21 February 2011. [Online]. Available: <http://www.cancernz.org.nz/divisions/auckland/about/cancer-statistics>. [Accessed 28 May 2014].
- [2] International Agency for Research on Cancer, "GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012," World Health Organization, 2012. [Online]. Available: [http://globocan.iarc.fr/Pages/fact\\_sheets\\_cancer.aspx](http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx). [Accessed 5 June 2014].
- [3] International Agency for Research on Cancer (IARC), *Global battle against cancer won't be won with treatment alone Effective prevention measures urgently needed to prevent cancer crisis*, Lyon/London: World Health Organization, 2014.
- [4] W. H. Organization, *National Cancer Control Programmes, Policies and managerial guidelines*, 2nd Edition, Geneva: World Health Organization, 2002.
- [5] S. J. C. F. M. B. Geoff Delaney, "The Role of Radiotherapy in Cancer Treatment," *Cancer*, vol. 104, no. 6, pp. 1129–1137, 2005.
- [6] National Cancer Institute, "Radiation Therapy for Cancer," National Institutes of Health, 30 06 2012. [Online]. Available: <http://www.cancer.gov/cancertopics/factsheet/Therapy/radiation#r1>. [Accessed 7 6 2014].

- [7] World Health Organization, "What is Ionizing Radiation?," World Health Organization, 2014. [Online]. Available: [http://www.who.int/ionizing\\_radiation/about/what\\_is\\_ir/en/](http://www.who.int/ionizing_radiation/about/what_is_ir/en/). [Accessed 29 7 2014].
- [8] F. H. Attix, Introduction to radiological physics and radiation dosimetry, Wiley-VCH, 1991.
- [9] M. C. Malley, Radioactivity: A History of a Mysterious Science, New York: Oxford University Press, 2011.
- [10] M. H. N. Conyers Herring, "Thermionic Emission," *Rev. Mod. Phys.*, vol. 21, no. 2, pp. 185–270, 1949.
- [11] A. v. d. K. Michael Joiner, Basic Clinical Radiobiology 4th edition, London: Hodder Arnold, 2009.
- [12] World Health Organization, "Cancer," World Health Organization, 2014. [Online]. Available: <http://www.who.int/topics/cancer/en/>. [Accessed 3 7 2014].
- [13] International Commission on Radiation Units and measurements (ICRU), "Prescribing, Recording and Reporting Photon Beam (Report 62) (Supplement to ICRU Report 50)," International Commission on Radiation Units and measurements (ICRU), Maryland, 1999.
- [14] S. J. T. E. B. J. J. Neil G Burnet, "Defining the tumour and target volumes for radiotherapy," *Cancer Imaging*, vol. 4, pp. 153–161, 2004.
- [15] A. T. G. 101, "Stereotactic body radiation therapy: The report of AAPM TaskGroup 101," *Medical Physics*, vol. 37, no. 8, pp. 4078–4101, 2010.
- [16] S. Meeks, "Immobilization from rigid to non-rigid," in *AAPM Summer School*, Burnaby, 2011.

- [17] L. A. e. a. Dawson, "The reproducibility of organ position using active breathing control (ABC) during liver radiotherapy," *International Journal of Radiation Oncology • Biology • Physics*, vol. 51, no. 5, pp. 1410–1421, 2000.
- [18] J. Talbot, *A Patient Position Guidance System in Radiotherapy Using Augmented Reality*, Christchurch: University of Canterbury.
- [19] B. Curless, "From Range Scans to 3D Models," *ACM SIGGRAPH Computer Graphics*, vol. 4, no. 33, p. 38–41, 2000.
- [20] P. Z. G. M. C. Carlo Dal Mutto, *Time-of-Flight Cameras and Microsoft Kinect*, New York: Springer, 2012.
- [21] S. Sumitro, "Application of Smart 3-D Laser Scanner in structural health monitoring," *Proceedings of SMSST'07, World Forum on Smart Materials and Smart Structures Technology (SMSST'07)*, vol. 568, 2007.
- [22] S. Schuon, C. Theobalt, J. Davis and S. Thrun, "High-quality scanning using time-of-flight depth superresolution," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, 2008.
- [23] Microsoft, "Kinect for Windows v2 SDK documentation," Microsoft, Seattle, 2013.
- [24] Adadruit Industries, "WE HAVE A WINNER – Open Kinect driver(s) released – Winner will use \$3k for more hacking – PLUS an additional \$2k goes to the EFF!," Adadruit Industries, 10 November 2010. [Online]. Available: <http://www.adafruit.com/blog/2010/11/10/we-have-a-winner-open-kinect-drivers-released-winner-will-use-3k-for-more-hacking-plus-an-additional-2k-goes-to-the-eff/>. [Accessed 28 February 2014].

- [25] H. L. D. Fukunaga Keinosuke, "The Estimation of the Gradient of a Density Function, with Application in Pattern Recognition," *IEEE Transaction on Information Theory*, vol. 21, no. 1, pp. 32–40, 1975.
- [26] Gary R. Bradski, Microcomputer Research Lab, Santa Clara, CA, Intel Corporation, "Computer Vision Face Tracking For Use in a Perceptual User Interface," *Intel Technology Journal*, no. Q2, pp. 1–15, 1998.
- [27] D. G. Lowe, "Object Recognition from local Scale-Invariant Features," University of British Columbia, Vancouver, 1999.
- [28] T. Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales," *Journal of Applied Statistics*, vol. 21, no. 2, pp. 224–270, 1994.
- [29] J. a. D. G. L. Beis, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," in *Conference on Computer Vision and Pattern Recognition*, Puerto Rico, 1997.
- [30] T. T. L. V. G. Herbert Bay, "SURF: Speeded Up Robust Features," *Computer Vision ECCV*, pp. 404–417, 2006.
- [31] N. D. M. Paul J Besl, "A Method for Registration of 3-D Shapes," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no. 2, pp. 239–256, 1992.
- [32] D. K. O. H. D. M. R. N. P. K. J. S. S. H. D. F. A. D. A. F. Shahram Izadi, "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera," in *ACM Symposium on User Interface Software and Technology*, Santa Barbara, 2011.



- [33] D. F. a. H. muller, "Impact of Thermal and Environmntal Conditions on the Kinect Sensor," *Advances in Depth Image Analysis and Applications. Springer Berlin Heidelberg*, pp. 21–31, 2013.
- [34] G. Wiora, "Laserprofilometer EN," 10 April 2006. [Online]. Available: [http://en.wikipedia.org/wiki/3D\\_scanner#mediaviewer/File:Laserprofilometer\\_EN.svg](http://en.wikipedia.org/wiki/3D_scanner#mediaviewer/File:Laserprofilometer_EN.svg). [Accessed 15 7 2014].
- [35] wildman, "Programming for Kinect 4 – Kinect App with Skeleton Tracking," 3dsense interactive technologies, 22 6 2013. [Online]. Available: <http://blog.3dsense.org/programming/programming-for-kinect-4-kinect-app-with-skeleton-tracking/>. [Accessed 24 7 2014].
- [36] tomq, "How to convert Depth cam data to world coordinates," as3NUI, 17 2 2012. [Online]. Available: <http://forum.as3nui.com/viewtopic.php?f=6&t=36>. [Accessed 24 7 2014].
- [37] L. F. S. Robert A Novelline, *Squire's Fundamentals of Radiology*, Cambridge: Harvard University Press, 1997.
- [38] H. T. A. U. F. K. Paul Milgram, "AUGMENTED REALITY: A CLASS OF DISPLAYS ON THE REALITY-VIRTUALITY CONTINUUM," *Telemanipulator and Telepresence Technologies*, vol. 2351, no. 34, pp. 282–292, 1994.
- [39] M. Musaddig, "LINEAR ACCELERATOR," 3 July 2009. [Online]. Available: <http://bmeng.blogspot.co.nz/2009/07/linear-accelerator-linac.html>. [Accessed 29 7 2014].
- [40] G. M. MacKee, *X-rays and radium in the treatment of diseases of the skin*, Lea & Febiger, 1921.

- [41] Koukalaka, "Physics in Medicine Week 3: Radiotherapy," 24 1 2012. [Online]. Available: <http://koukalaka.wordpress.com/2012/01/24/physics-in-medicine-week-3-radiotherapy/>. [Accessed 29 7 2014].
- [42] J. Kelly, "Production of x-rays for Diagnosis," University College London, 1995. [Online]. Available: <http://img.chem.ucl.ac.uk/www/kelly/medicalxrays.htm>. [Accessed 17 7 2014].
- [43] H. Kato, "Computer Vision Algorithm," Human Interface Technology Laboratory, [Online]. Available: <http://www.hitl.washington.edu/artoolkit/documentation/vision.htm>. [Accessed 17 06 2014].
- [44] John, "Matlab Answers," 10 May 2013. [Online]. Available: <http://www.mathworks.com/matlabcentral/answers/75287-does-anyone-do-research-on-structured-light-how-to-get-a-good-edge-detection>. [Accessed 15 7 2014].
- [45] J. Hykes, Artist, *Pb-gamma-xs*. [Art]. 2011.
- [46] P. Frame, "Coolidge X-ray Tubes," Oak Ridge Associated Universities, 24 06 2009. [Online]. Available: <https://www.ornl.gov/ptp/collection/xraytubescoolidge/coolidgeinformation.htm>. [Accessed 2 07 2014].
- [47] S. Clarkson, "Science Behind 3D Vision," Centre for Sports Engineering Research, Sheffield Hallam University, [Online]. Available: <http://www.depthbiomechanics.co.uk/?p=102>. [Accessed 24 7 2014].
- [48] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 17, no. 8, pp. 790-799, 1995.
- [49] J. Belot, Radiotherapy in skin disease, General Books LLC, 1905.

- [50] admin, "Kinect Hacking 103: Looking at Kinect IR Patterns.," Futurepicture, 17 11 2010. [Online]. Available: <http://www.futurepicture.org/?p=116>. [Accessed 15 7 2014].
- [51] Sutter Health, "RapidArc Radiotherapy," Alta Bates Summit Medical Center, 2013. [Online]. Available: <http://www.altabatesummit.org/rapidarc/>. [Accessed 07 06 2014].
- [52] Aktina medical, "Model LINAC, IEC 60601 Scales," Model LINAC, IEC 60601 Scales, 2013. [Online]. Available: <http://www.aktina.com/product/model-linac-iec-60601-scales/>. [Accessed 17 7 2014].
- [53] Microsoft , "Kinect," Microsoft, [Online]. Available: <http://www.xbox.com/en-US/xbox-one/meet-xbox-one?xr=shellnav#adrenalinejunkie>. [Accessed 15 7 2014].
- [54] Varian Medical Systems, "Image Gallery: Multileaf Collimators," Varian medical systems, 2014. [Online]. Available: <http://newsroom.varian.com/index.php?s=31899&mode=gallery&cat=2473>. [Accessed 7 06 2014].
- [55] Southwest research institute, "Current Biometrics Device Development Initiatives," Southwest research institute, 15 4 2014. [Online]. Available: <http://www.swri.org/4org/d14/ElectroSys/biometrics/current.htm>. [Accessed 15 7 2014].
- [56] Amptek, "CdTe Application Note: Characterization of X-ray Tubes," Amptek, 2014. [Online]. Available: <http://www.amptek.com/cdte-application-note-characterization-of-x-ray-tubes/>. [Accessed 2 7 2014].

[57] Teach Nuclear, "Cancer Theapy," Teach Nuclear, 2014. [Online]. Available: [http://teachnuclear.ca/contents/cna\\_nuc\\_tech/med\\_app\\_intro/cobalt60/](http://teachnuclear.ca/contents/cna_nuc_tech/med_app_intro/cobalt60/). [Accessed 1 07 2014].